Rewolucyjna... jedna z najbardziej mentalnie zwinnych, intelektualnie pomysłowych książek jakie masz szansę przeczytać. — „GUARDIAN"

Daniel C. Dennett

ŚWIADOMOŚĆ

Daniel C. Dennett

ŚWIADOMOŚĆ

Tłumaczenie
EWA STOKŁOSA

Redakcja naukowa i Posłowie:
MARCIN MIŁKOWSKI

Rysunki:
PAUL WEINER

Copernicus Center PRESS

Przedmowa

Na pierwszym roku college'u przeczytałem *Medytacje* Kartezjusza i zainteresował mnie problem umysł-ciało. Oto tajemnica. Jak, do diabła, moje myśli i odczucia mogą mieścić się w tym samym świecie co komórki nerwowe i cząsteczki tworzące mój mózg? Dziś, po trzydziestu latach rozmyślania, rozmów i pisania o tej tajemnicy, sądzę, że zrobiłem postęp. Wydaje mi się, że potrafię naszkicować schemat rozwiązania, teorię świadomości, dającą odpowiedzi (lub pokazującą, jak je znaleźć) na pytania, które dotychczas były tak samo zaskakujące dla filozofów i naukowców, jak dla laików. Otrzymałem dużą pomoc. Miałem szczęście być nauczanym, nieformalnie, niestrudzenie i nieustępliwie, przez wspaniałych myślicieli, których poznacie na stronach tej książki. Historia, którą muszę opowiedzieć, nie jest historią samotnego zamyślenia, lecz odyseją przez wiele dziedzin, a rozwiązania zagadek są nierozerwalnie wplecione w tkaninę dialogu i sporów, w których często więcej nauczą nas szalone błędy niż ostrożne ekwiwokacje. Jestem pewien, że w przedstawionej tu teorii pozostało wiele błędów, i mam nadzieję, iż są one właśnie szalone, gdyż sprowokują u innych lepsze odpowiedzi.

Ideom z tej książki nadawałem kształt przez szereg lat, jednak pisanie rozpocząłem w styczniu 1990 roku, a skończyłem ledwie rok później, dzięki hojności kilku wspaniałych instytucji i pomocy wielu przyjaciół, studentów i współpracowników. Zentrum für Interdisziplinäre Forschung w Bielefeld, CREA na École Polytechnique w Paryżu i Villa Serbelloni Rockefeller Center w Bellagio zapewniły idealne warunki do pisania i dyskusji podczas pierwszych pięciu miesięcy. Moja uczelnia, Tufts, wspierała mą pracę w Centrum Studiów Kognitywnych i umożliwiła mi zaprezentowanie prawie ostatniego szkicu jesienią 1990 roku podczas seminarium organizowanego dla Uniwersytetu Tufts oraz innych doskonałych szkół z Bostonu i okolic. Chciałbym też podziękować Fundacji Kapor za wspieranie naszych badań w Centrum Badań Kognitywnych.

Kilka lat temu Nicholas Humphrey przyłączył się do mnie w pracy w Centrum Badań Kognitywnych i wraz z Rayem Jackendoffem i Marcelem Kinsbourne'em rozpoczęliśmy regularne dyskusje nad różnymi aspektami i problemami świadomości. Trudno byłoby znaleźć cztery bardziej różniące się podejścia do umysłu, jednak nasze rozmowy były na tyle owocne i krzepiące, że dedykuję tę książkę tym wspaniałym przyjaciołom, dziękując im za to wszystko, czego mnie nauczyli. Dwaj inni długoletni przyjaciele również odegrali znaczące role w kształtowaniu mojego myślenia, za co jestem im dozgonnie wdzięczny: Kathleen Akins i Bo Dahlbom.

Pragnę także podziękować grupie ZIF z Bielefeld, a szczególnie takim osobom jak: Peter Bieri, Jaegwon Kim, David Rosenthal, Jay Rosenberg, Eckhart Scheerer, Bob van Gulick, Hans Flohr oraz Lex van der Heiden; grupie CREA z Paryża, a zwłaszcza Danielowi Andlerowi, Pierre'owi Jacobowi, Franciscowi Vareli, Danowi Sperberowi i Deirdre Wilson; oraz „księciom świadomości", którzy dołączyli do Nicka, Marcela, Raya i mnie w Villi Serbelloni podczas niesamowicie produktywnego tygodnia w marcu. Byli to: Edoardo Bisiach, Bill Calvin, Tony Marcel i Aaron Sloman. Dziękuję również Edoardowi i innym uczestnikom warsztatów dotyczących zaniedbania, a odbywających się w Parmie w czerwcu. Pim Levelt, Odmar Neumann, Marvin Minsky, Oliver Selfridge i Nils Nilsson też dostarczyli cennych rad dotyczących różnych rozdziałów. Chcę ponadto wyrazić swoją wdzięczność Nilsowi za dostarczenie fotografii Shakeya oraz Paulowi Bach-y-Rita za jego zdjęcia i porady dotyczące

urządzeń protezy wzroku.

Jestem wdzięczny za ogrom konstruktywnej krytyki wszystkim uczestnikom jesiennego seminarium, klasie, której nigdy nie zapomnę. Byli to: David Hilbert, Krista Lawlor, David Joslin, Cynthia Schossberger, Luc Faucher, Steve Weinstein, Oakes Spalding, Mini Jaikumar, Leah Steinberg, Jane Andreson, Jim Beattie, Evan Thompson, Turhan Canli, Michael Anthony, Martina Roepke, Beth Sangree, Ned Block, Jeff McConnell, Bjørn Ramberg, Phil Holcomb, Steve White, Owen Flanagan i Andrew Woodfield. Tydzień po tygodniu, ta grupa wystawiała mnie na próbę ognia w sposób najbardziej konstruktywny z możliwych. Podczas ostatecznego redagowania Kathleen Akins, Bo Dahlbom, Doug Hofstadter i Sue Stafford dostarczyli mi wiele nieocenionych sugestii. Paul Weiner zamienił moje surowe szkice we wspaniałe ryciny i wykresy.

Kathryn Wynes, a później Anne Van Voorhis wykonały nadzwyczajną pracę, powstrzymując mnie i Centrum od rozsypania się podczas ostatnich gorączkowych lat, a bez ich skuteczności i spostrzeżeń dokończenie tej książki nadal wymagałoby lat. Jednak największe podziękowania należą się Susan, Peterowi, Andrei, Marvinowi i Brandonowi, mojej rodzinie.

Uniwersytet Tufts

styczeń 1991

Rozdział 1

Preludium: Jak możliwe są halucynacje?

1. Mózg w naczyniu

Wyobraź sobie, że w czasie twojego snu, zły naukowiec wyciągnął twój mózg z ciała i włożył go do podtrzymującego życie naczynia. Wyobraź sobie również, że naukowiec sprawił, iż nie wydaje ci się, że jesteś tylko mózgiem w naczyniu, ale że działasz i zajmujesz się wieloma przyziemnymi sprawami, wymagającymi ciała. Ta stara historyjka, mózg w naczyniu, to ulubiony eksperyment myślowy w warsztacie wielu filozofów. To współczesna wersja opowieści Kartezjusza (1641) o złośliwym demonie-iluzjoniście, który pragnie oszukać Kartezjusza co do wszystkiego, łącznie z jego własną egzystencją. Ale, jak zaobserwował Kartezjusz, nawet nieskończenie potężny zły demon nie byłby w stanie nabrać go tak, aby ten myślał, że istnieje, gdyby w rzeczywistości nie istniał: *cogito ergo sum*, „myślę, więc jestem". Współczesnych filozofów mniej interesuje udowadnianie czyjejś egzystencji jako myślącego bytu (być może dlatego, że stwierdzili, iż Kartezjusz dostatecznie rozwiązał ten problem), a bardziej skupiają się na tym, co możemy wywnioskować z naszego doświadczenia z naturą oraz z samej natury świata, w którym (pozornie) żyjemy. Czy można być jedynie mózgiem w naczyniu? Czy można zawsze być tylko mózgiem w naczyniu? Jeśli tak, czy można zdać sobie z tego sprawę (a co dopiero to zweryfikować)?

Pomysł z mózgiem w naczyniu doskonale nadaje się do zgłębiania tych zagadnień, ale ja chciałbym użyć tej historyjki w trochę innym celu. Zamierzam przedstawić kilka ciekawych faktów dotyczących halucynacji, które doprowadzą nas do zalążka teorii – empirycznej, rzetelnej teorii – ludzkiej świadomości. W standardowym eksperymencie myślowym naukowcy oczywiście mieliby pełne ręce roboty, próbując zapewnić kikutom nerwów wszystkich zmysłów odpowiednią stymulację, aby podstęp mógł dojść do skutku, ale filozofowie dla dobra dyskusji założyli, że pomimo trudności, które wiążą się z tym zadaniem, jest ono „zasadniczo możliwe". Należy być nieufnym co do rzeczy „zasadniczo możliwych". Zasadniczo możliwe jest również wybudowanie drabiny ze stali nierdzewnej prowadzącej do nieba lub wypisanie w porządku alfabetycznym wszystkich możliwych w języku angielskim rozmów składających się z mniej niż tysiąca słów. Żaden jednak z tych pomysłów nie jest w minimalnym stopniu możliwy do zrealizowania, a – jak zobaczymy – czasem *faktyczna niemożliwość* jest teoretycznie ciekawsza niż *zasadnicza możliwość*.

Zastanówmy się przez chwilę, jak zatrważające zadanie miałby przed sobą naukowiec. Można przypuszczać, że po łatwym początku pojawiłyby się trudności. Naukowiec zacząłby od mózgu w stanie śpiączki, który byłby podtrzymywany przy życiu, ale nie otrzymywałby sygnałów od nerwów wzrokowych czy słuchowych, bodźców somatosensorycznych ani żadnych innych impulsów. Czasem przypuszcza się, że taki „odłączony" mózg pozostałby na zawsze w stanie śpiączki, nie potrzebując morfiny do uśpienia, lecz istnieją dowody na to, iż w tak strasznych warunkach mogłoby nastąpić spontaniczne wybudzenie. Myślę, iż można założyć, że gdyby ktoś miał wybudzić się w takiej sytuacji, znalazłby się w stanie okrutnej udręki: niewidomy, niesłyszący, całkowicie sparaliżowany, bez możliwości określenia orientacji swojego ciała.

Gdyby naukowcy nie chcieli cię przerazić, próbowaliby wybudzić cię poprzez wprowadzenie muzyki stereo (w odpowiedni sposób zakodowanej jako impulsy nerwowe) do twoich nerwów słuchowych. Zaaranżowaliby również sygnały, zwykle pochodzące od twojego zmysłu równowagi, czyli z ucha wewnętrznego, które wskazałyby ci, że leżysz na plecach, ale sparaliżowany, otępiały, niewidomy. Taki stan będzie do osiągnięcia poprzez techniczną wirtuozerię w najbliższej przyszłości – być może jest dostępny już dziś. Następnie naukowcy mogliby przejść do stymulowania ośrodków unerwiających twój naskórek, zapewniając im sygnały, które w normalnej sytuacji mogłyby zostać wysłane z pomocą delikatnego ciepła do powierzchni brzusznej, oraz (idąc jeszcze dalej) zacząć stymulować nerwy naskórka na grzbiecie, wywołując mrowienie charakterystyczne dla ziarenek piasku wciskających się w twoje plecy. „Świetnie!", powiedziałbyś do siebie. „Leżę sobie tutaj na plaży, sparaliżowany i niewidomy, słuchając przyjemnej muzyki, ale zagrożony poparzeniem słonecznym. Jak się tu dostałem i jak mogę zawołać pomoc?"

Następnie wyobraź sobie, że osiągnąwszy to wszystko, naukowcy stawiają czoło trudniejszemu problemowi, polegającemu na przekonaniu cię, że nie jesteś zwykłą kłodą leżącą na plaży, lecz podmiotem działającym. Zaczynając od małych kroczków, postanawiają „zlikwidować" część paraliżu w twoim fantomowym ciele i pozwalają ci poruszyć prawym palcem wskazującym w piasku. Pozwalają na zmysłowe odczucie poruszanego palca, co jest możliwe poprzez przekazanie kinestetycznej odpowiedzi, związanej z odpowiednimi sygnałami motorycznymi w obwodowym układzie nerwowym, ale muszą również zlikwidować odrętwienie z fantomu oraz zapewnić stymulację dla odczucia, które mógłby wywołać piasek wokół twojego palca.

Nagle zdajemy sobie sprawę z problemu, który szybko wymknie się spod kontroli, a mianowicie, że to, jak poczujesz piasek, zależy od sposobu, w jaki postanowisz poruszyć palcem. Problem związany z oszacowaniem odpowiednich doznań, ich wygenerowaniem lub skomponowaniem, a następnie z zaprezentowaniem ich w czasie rzeczywistym, będzie matematycznie nie do rozwiązania przez żaden, nawet najszybszy komputer, a jeśli źli naukowcy postanowią rozwiązać problem czasu rzeczywistego, wcześniej obliczając i przechowując wszelkie możliwe reakcje i w odpowiednim momencie puszczając je z playbacku, pojawi się kolejny nierozwiązywalny problem: do przechowania będą mieli zbyt wiele możliwości. Znaczy to po prostu, że naukowy zostaną zalani *eksplozją kombinatoryczną*[1], gdy tylko dadzą ci jakiekolwiek autentyczne możliwości eksploracyjne w tym wyimaginowanym świecie.

Naukowcy natrafili na znajomy mur; jego cień widzimy w nudnych stereotypach każdej gry komputerowej. Możliwości działania muszą być surowo – i nierealistycznie – ograniczone, aby zadanie stojące przed twórcami światów było wykonalne. Jeśli naukowców stać jedynie na to, aby przekonać cię, że jesteś skazany na granie w *Donkey Konga* przez całe twoje życie, znaczy to, że są naprawdę złymi naukowcami.

Istnieje pewne rozwiązanie tego technicznego problemu. Jest to rozwiązanie wykorzystywane na przykład w sytuacji, gdy potrzeba ulżyć obliczeniowemu ciężarowi w wysoce realistycznych symulatorach lotu: użycie *kopii* rzeczy w symulowanym świecie. Używają prawdziwego kokpitu, którym się porusza za pomocą popychacza hydraulicznego, zamiast symulować te odczucia w spodniach trenującego pilota. Oznacza to po prostu, że istnieje tylko jeden sposób na przechowywanie tak ogromnej ilości informacji o gotowym dostępie, dotyczących wyimaginowanego świata do eksploracji, a jest to *prawdziwy* (nawet jeśli maleńki, sztuczny albo gipsowy) świat, który przechowuje informacje o sobie. Jest to „oszukiwanie", jeśli jesteś złym demonem, twierdzącym, iż oszukał Kartezjusza co do istnienia absolutnie wszystkiego, ale jest to sposób rozwiązania problemu z użyciem, bądź co bądź, ograniczonych

środków.

Kartezjusz postąpił mądrze, obdarzając swojego wyimaginowanego złego demona *nieskończoną* mocą podstępu. Mimo że zadanie nie jest, ściśle mówiąc, nieskończone, ilość informacji możliwych do uzyskania w krótkim czasie przez dociekliwego człowieka jest oszałamiająco ogromna. Inżynierowie mierzą przepływ informacji w bitach na sekundę, czyli mówią o szerokości pasma kanałów, którymi płyną informacje. Telewizja wymaga szerszego pasma niż radio, a telewizja wysokiej rozdzielczości pasma jeszcze szerszego. Telewizja wysokiej rozdzielczości wydzielająca zapachy i produkująca fizyczne odczucia wymagałaby pasma jeszcze szerszego, a telewizja taka sama, tylko interaktywna, potrzebowałaby astronomicznie szerokiego pasma, ponieważ nieustannie rozgałęziałaby się na tysiące, odrobinę różniących się od siebie, trajektorii w wyimaginowanym świecie. Daj sceptykowi wątpliwą monetę, a w dwie sekundy ważenia, drapania, dzwonienia, nadgryzania i zwykłego patrzenia, jak słońce odbija się od jej powierzchni, zbierze on więcej informacji, niż superkomputer Cray jest w stanie wygenerować w rok. Wybicie *rzeczywistej*, ale podrobionej monety to pestka; stworzenie *symulowanej* monety jedynie ze zorganizowanej symulacji nerwów leży poza możliwościami ludzkiej technologii, współczesnej i prawdopodobnie przyszłej[2].

Stąd płynie pierwszy wniosek: nie jesteśmy mózgami w naczyniach – jeśli cię to martwiło. Kolejny wniosek jest taki, że silne halucynacje są po prostu niemożliwe! Przez silną halucynację rozumiem halucynację pozornie konkretnego i trwałego trójwymiarowego przedmiotu w świecie rzeczywistym – w odróżnieniu od błysków, zniekształceń geometrycznych, aur, powidoków, nagłych odczuć fantomowych kończyn i innych anomalnych wrażeń. Silną halucynacją mógłby być na przykład duch, który odpowiada, pozwala się dotknąć i sprawia wrażenie ciała stałego, rzuca cień i jest widoczny pod każdym kątem tak, że można go okrążyć i zobaczyć, jak wygląda z tyłu.

Halucynacje mogą być sklasyfikowane pod względem siły oddziaływania poprzez kilka tego rodzaju cech. Relacje z przeżywania silnych halucynacji są rzadkie i widzimy teraz, dlaczego nie jest przypadkiem, że ich wiarygodność intuicyjnie wydaje się odwrotnie proporcjonalna do siły zrelacjonowanej halucynacji. Jesteśmy – i powinniśmy być – wyjątkowo sceptyczni co do relacji o silnych halucynacjach, ponieważ nie wierzymy w duchy, a uważamy, że tylko prawdziwy duch mógłby wywołać silną halucynację. (Siła halucynacji zrelacjonowanych przez Carlosa Castañedę w *Naukach Don Juana* [1968/1991] była pierwszym czynnikiem, który skłonił naukowców do przypuszczenia, że książka jest fikcją, nie faktem, mimo iż autor otrzymał doktorat z antropologii na uniwersytecie UCLA za swoje „badania" nad Don Juanem).

Nie znamy przypadków bardzo silnych halucynacji, jednak bez wątpienia często zdarzają się przekonujące halucynacje związane z kilkoma modalnościami sensorycznymi. Halucynacje dobrze potwierdzone w literaturze psychologii klinicznej są często szczegółowymi fantazjami, które wychodzą daleko poza wytwórcze możliwości współczesnej technologii. W jaki sposób jeden mózg jest w stanie dokonać tego, co jest praktycznie niemożliwe dla zastępów naukowców i animatorów komputerowych? Jeśli takie przeżycie nie jest autentycznym czy prawdziwym postrzeganiem czegoś rzeczywistego „poza" umysłem, musi ono być całkowicie wytworzone w umyśle (lub w mózgu), wyssane z palca, a jednocześnie wystarczająco realistyczne, żeby oszukać umysł, który je preparuje.

2. Dowcipnisie w mózgu

O halucynacjach przypuszcza się najczęściej, że pojawiają się, kiedy w mózgu dochodzi do jakiejś dziwnej autostymulacji, a dokładniej rzecz ujmując, całkowicie wewnętrznie

wytworzonej stymulacji pewnych części lub poziomów systemów percepcyjnych mózgu. Kartezjusz w XVII wieku dostrzegł to wyraźnie, gdy pisał o kończynach fantomowych, co jest zaskakującą, choć zupełnie normalną halucynacją, w której po amputowaniu kończyny pacjent zdaje się odczuwać nie tylko obecność odciętej części ciała, ale również jej swędzenie, mrowienie czy ból. (Wrażenia ciągłej obecności kończyny są tak żywe i realistyczne, że często zdarza się, iż pacjenci po amputacji po prostu nie są w stanie uwierzyć, że noga czy stopa zostały odcięte, do momentu, w którym widzą ich brak na własne oczy). Analogią Kartezjusza były dzwonki dla służby. Zanim wynaleziono dzwonki elektryczne, domofony czy walkie-talkie, wielkie domy były wyposażone we wspaniałe systemy drutów i kół linowych, które pozwalały na wezwanie służby z każdego pomieszczenia w domu. Silne szarpnięcie aksamitnej szarfy zwisającej z dziury w ścianie pociągało drut, który prowadził po kołach linowych aż do spiżarni, informując lokaja, że wymagana jest obsługa w głównej sypialni, w salonie lub w pokoju bilardowym. System funkcjonował dobrze, ale był pożywką dla dowcipnisiów. Pociągnięcie za drut na jakiejkolwiek jego długości sprawiało, że lokaj pędził do salonu, głęboko wierząc, iż ktoś go tam wezwał – skromny rodzaj halucynacji. Kartezjusz uważał, że podobnie jest z percepcją, powodowaną różnymi skomplikowanymi ciągami zdarzeń w układzie nerwowym, które prowadzą w końcu do centrum dowodzenia świadomego umysłu, i że gdyby ktoś był w stanie interweniować w którymkolwiek miejscu tego łańcucha (na przykład na jakiejkolwiek długości nerwu wzrokowego między okiem a świadomością), pociągnięcie za nerwy doprowadziłoby do dokładnie takiego samego łańcucha zdarzeń, jaki powstałby w normalnym, autentycznym postrzeganiu czegoś, oraz spowodowałoby w odbiorczej części umysłu dokładnie taki sam efekt świadomej percepcji.

Mózg lub jego część niechcący spłatał umysłowi mechanicznego figla. Takie było Kartezjańskie wyjaśnienie halucynacji kończyn fantomowych. Halucynacje te, mimo że są niezwykle żywe, w naszej terminologii są raczej halucynacjami słabymi; składają się z niezorganizowanego bólu i swędzenia, należących do jednej modalności sensorycznej. Pacjenci po amputacji nie widzą, nie słyszą ani (o ile wiem) nie czują zapachu swoich fantomowych stóp. Tak więc ujęcie Kartezjusza mogłoby być poprawnym wyjaśnieniem zjawiska fantomowych kończyn, jeśli na razie nie weźmiemy pod uwagę znanej tajemnicy dotyczącej interakcji fizycznego mózgu z niefizycznym świadomym umysłem. Widzimy jednak, że nawet czysto mechaniczna strona historii Kartezjusza musi być błędna, gdy mowa jest o stosunkowo silnych halucynacjach; mózg jako iluzjonista w żaden sposób nie byłby w stanie przechowywać wystarczającej ilości fałszywych informacji i manipulować nimi tak, aby nabrać dociekliwy umysł. Mózg może się zrelaksować i pozwolić rzeczywistemu światu dostarczać nadmiar prawdziwych informacji, ale jeśli zacznie doprowadzać do spięcia w swoich własnych nerwach (lub ciągnąć za swoje własne druty, jak powiedziałby Kartezjusz), rezultatem będzie jedynie najsłabsza z ulotnych halucynacji. (Podobnie awaria suszarki elektrycznej sąsiada może wywołać „zaśnieżony ekran" lub szum bądź brzęczenie, czy wręcz dziwne błyski na ekranie mojego telewizora, ale jeśli obejrzę fałszywą wersję wieczornych wiadomości, będę *wiedział*, że przyczyną jest wysoce dopracowana organizacja, która wychodzi daleko poza talenty suszarki do włosów).

Kuszące jest przypuszczenie, że być może jesteśmy zbyt łatwowierni w kwestii halucynacji; niewykluczone, że zachodzą tylko umiarkowane, ulotne, kiepskie halucynacje – silne nie wydarzają się, ponieważ nie mogą się wydarzyć! Pobieżny przegląd literatury dotyczącej halucynacji zdecydowanie wskazuje na to, że ich siła i częstotliwość są odwrotnie proporcjonalne – tak jak siła i wiarygodność. Daje nam jednak również wskazówkę prowadzącą do kolejnej teorii mechanizmów wytwarzających to zjawisko: jedną z charakterystycznych cech

doniesień o halucynacji jest fakt, że ofiara przyznaje się do swojej dosyć niezwykłej bierności w obliczu tego przeżycia. Osoby mające halucynacje zwykle po prostu stoją i są zdumione. Zazwyczaj nie mają potrzeby jej zbadania, podważenia czy zakwestionowania i nie podejmują żadnych kroków, aby wejść w interakcję ze zjawą. Niewykluczone, z powodów, które właśnie zbadaliśmy, że ta bierność nie jest nieistotną cechą halucynacji, lecz koniecznym warunkiem powstania umiarkowanie szczegółowej i trwałej ułudy.

Natomiast tylko w razie bierności ze strony ofiary silne halucynacje mogłyby przetrwać. Powód tego jest taki, że iluzjonista – czyli to, co wytwarza halucynację – może „liczyć" na konkretny sposób eksploracji przez ofiarę – w przypadku absolutnej bierności jest to *żaden* rodzaj eksploracji. Dopóki iluzjonista może szczegółowo przewidzieć sposób eksploracji, który zostanie podjęty, dopóty przygotowuje się na to, że iluzja będzie musiała być podtrzymana „w kierunkach, w które spojrzy ofiara". Projektanci planów filmowych z góry wymagają informacji o ulokowaniu kamery – a jeśli nie będzie stacjonarna, to informacji o jej trajektorii i kącie – ponieważ wtedy muszą przygotować tylko tyle materiałów, aby pokryć filmowaną perspektywę. (Nie bez powodu twórcy *cinéma vérité* często filmują z ręki, która swobodnie wędruje po planie). Ta sama zasada była używana przez Potiomkina, aby zaoszczędzić na wioskach pokazywanych Katarzynie Wielkiej; plan jej podróży musiał być ściśle przestrzegany.

Tak więc pewnym rozwiązaniem problemu silnych halucynacji jest założenie, iż istnieje połączenie między ofiarą i iluzjonistą, które umożliwia iluzjoniście zbudowanie iluzji *zależnej* od ofiary, a zatem mogącej przewidywać jej eksploracyjne zamierzenia i decyzje. Gdy iluzjonista nie jest w stanie „czytać w myślach" ofiary, aby wyciągnąć z nich informacje, może czasem w prawdziwym życiu (na przykład magik na scenie) *wprowadzić* pewną ścieżkę dociekań poprzez subtelne, ale silne „wmuszanie psychologiczne". I tak magik karciany zna wiele standardowych sposobów wmówienia ofierze, że wypełniana jest jej wolna wola co do tego, jakie karty na stole zbada, podczas gdy w rzeczywistości jest tylko jedna karta, która może zostać przez nią odwrócona. Powracając do naszego poprzedniego eksperymentu myślowego, powiemy, że jeśli zły naukowiec może *zmusić* mózg w naczyniu do tego, by miał konkretny zestaw zamierzeń eksploracyjnych, może rozwiązać problem eksplozji kombinatorycznej, przygotowując tylko przewidziany materiał; system będzie wówczas tylko *wydawał się* interaktywny. Podobnie zły demon Kartezjusza może podtrzymać iluzję, nie potrzebując nieskończonej siły, jeśli podtrzyma u ofiary złudzenie wolnej woli, gdyż badanie przez nią wyimaginowanego świata drobiazgowo kontroluje[3].

Istnieje natomiast jeszcze bardziej oszczędny (i realistyczny) tryb tworzenia halucynacji w mózgu, tryb, który ujarzmia rozbuchaną ciekawość ofiary. Możemy zrozumieć, jak on działa, przez analogię z pewną grą.

3. Gra zwana psychoanalizą

Gra polega na tym, że jakiejś osobie, naiwniakowi, mówi się, że gdy wyjdzie z pokoju, jedna z obecnych osób zostanie poproszona o zrelacjonowanie niedawnego snu. Wszyscy obecni poznają fabułę snu, a kiedy naiwniak powróci do pokoju i zacznie zadawać obecnym pytania, tożsamość osoby, której przyśnił się sen, nie zostanie przez nikogo ujawniona. Zadaniem naiwniaka będzie zadawać grupie pytania, na które można odpowiedzieć jedynie „tak" lub „nie" do czasu, gdy domyśli się wystarczająco szczegółowo, jak przebiegała narracja, a wtedy przeprowadzi psychoanalizę osoby, której przyśnił się sen, i na tej podstawie zidentyfikuje tę osobę.

Gdy naiwniak wychodzi z pokoju, prowadzący mówi obecnym, że nikt nie będzie

relacjonował swojego snu, mają jedynie odpowiadać na pytania według jednej prostej zasady: jeśli ostatnia litera ostatniego słowa w pytaniu znajduje się w pierwszej połowie alfabetu, na pytania będą odpowiadać twierdząco, na wszystkie inne pytania będą odpowiadać przecząco, z jednym zastrzeżeniem: istnieje zasada niesprzeczności, która jest ważniejsza od poprzedniej, mówiąca, iż odpowiedzi na późniejsze pytania nie mogą być sprzeczne z poprzednimi odpowiedziami. Na przykład:

Pytanie: Czy sen był o dziewczynie?

Odpowiedź: Tak.

Lecz jeśli później zapominalski naiwniak zapyta:

P: Czy we śnie pojawiły się jakieś dziewczyny lub kobiety?

O: Tak [pomimo litery *y* na końcu pytania stosujemy zasadę niesprzeczności][4].

Kiedy naiwniak wróci do pokoju i zacznie zadawać pytania, otrzyma mniej więcej losową, a w każdym razie przypadkową serię odpowiedzi pozytywnych i negatywnych. Rezultaty często okazują się zabawne. Czasami zabawa szybko kończy się absurdem, jak w przypadku pierwszego pytania „Czy fabuła snu jest identyczna co do szczegółu z fabułą książki *Wojna i pokój*?", albo „Czy we śnie pojawiły się jakiekolwiek istoty?". Częstszy rezultat to przedziwna i nierzadko obsceniczna historia o niedorzecznym nieszczęśliwym wypadku, ku uciesze uczestników. Kiedy naiwniak w końcu oświadcza, że osoba, której przyśnił się sen, kimkolwiek jest, musi być bardzo chorym i zaburzonym osobnikiem, uczestnicy z radością zawiadamiają naiwniaka, że on sam jest autorem „snu". To oczywiście nie do końca prawda. W pewnym sensie naiwniak jest autorem z racji pytań, które zadał. (Nikt inny nie zaproponował umieszczenia trzech goryli w łódce z siostrą zakonną). Z drugiej strony sen po prostu nie ma autora i w tym tkwi sedno. Widzimy tu proces narracyjnej produkcji, szczegółowego nagromadzenia, bez żadnych planów autora – iluzja bez iluzjonisty.

Struktura tej gry jest uderzająco podobna do struktury rodziny uznanych modeli systemów percepcyjnych. Powszechnie uważa się na przykład, że widzenie u człowieka nie może być wyjaśnione jako proces *wyłącznie* „sterowany danymi" lub „oddolny", ale na najwyższych poziomach musi być uzupełniony o kilka rund testowania hipotez „sterowanych oczekiwaniami" (lub czymś analogicznym). Do tego rodzaju modeli należy też model percepcji „analiza przez syntezę", który również zakłada, że postrzeganie powstaje w procesie wielokrotnego uzgadniania wytwarzanych oczekiwań z jednej strony i potwierdzeń (lub obaleń) powstających na peryferiach z drugiej strony (np. Neisser 1967). W tych teoriach chodzi przede wszystkim o to, że po iluś etapach „oddolnego przetwarzania" we wczesnych lub peryferyjnych warstwach systemu percepcyjnego postrzeganie kończy się dzięki cyklom generowania i testowania – przedmioty zostają zidentyfikowane, rozpoznane i skategoryzowane. W takim cyklu bieżące oczekiwania i zainteresowania tworzą hipotezę dla systemów percepcyjnych osobnika, które ją potwierdzają bądź obalają, a szybkie sekwencje generowania i potwierdzania takich hipotez prowadzą do ostatecznego wytworu – bieżącego, uaktualnionego „modelu" świata odbiorcy. Takie ujęcia percepcji są uzasadnione przez wiele względów biologicznych i epistemologicznych, a mimo że nie powiedziałbym, iż prawdziwość któregokolwiek z tych modeli została ostatecznie dowiedziona, to eksperymenty inspirowane tym podejściem dają niezłe rezultaty. Niektórzy teoretycy wręcz odważnie twierdzą, że percepcja *musi* mieć taką fundamentalną strukturę.

Bez względu na to, jaki będzie ostateczny werdykt w sprawie teorii generowania i testowania w percepcji, widzimy, że teoria ta uzasadnia proste i kompleksowe ujęcie halucynacji. Aby halucynacje pojawiły się w skądinąd normalnym systemie percepcyjnym, cykl generujący hipotezy (część sterowana oczekiwaniami) musi działać normalnie, a cykl sterowany danymi (czyli potwierdzający lub obalający) – funkcjonować w sposób zakłócony lub

przypadkowy, tak jak w grze towarzyskiej. Innymi słowy, jeśli szum w kanale informacyjnym zostanie przypadkowo wzmocniony jako „potwierdzenie" lub „obalenie" (przypadkowe odpowiedzi „tak" i „nie" w grze), to bieżące oczekiwania, zainteresowania, obsesje i troski ofiary będą prowadzić do tworzenia pytań lub hipotez, których treść jest odbiciem tych interesów ofiary, przez co „opowieść" powstanie w systemie percepcyjnym bez udziału autora. Nie musimy zakładać, że opowieść powstaje z wyprzedzeniem; nie musimy zakładać, że informacja jest przechowywana lub tworzona w części mózgu należącej do iluzjonisty. Zakładamy tylko, że iluzjonista przechodzi w tryb przypadkowego potwierdzania, a ofiara zapewnia treść, zadając pytania.

Ujęcie to wskazuje na bezpośredni związek między stanem emocjonalnym halucynującego a treścią tworzonych halucynacji. Halucynacje są zwykle treściowo powiązane z bieżącymi troskami osoby ich doświadczającej, a przedstawiony model halucynacji uwzględnia tę cechę bez potrzeby interwencji mało wiarygodnego wewnętrznego bajkopisarza, który miałby znać teorię modelu psychologii ofiary. Dlaczego na przykład myśliwy w ostatni dzień sezonu łowieckiego widzi jelenia z porożem i białym ogonem, gdy patrzy na czarną krowę lub na innego myśliwego w pomarańczowej kurtce? Ponieważ jego wewnętrzny osobnik zadający pytania naciska: „Czy to jeleń?" i otrzymuje odpowiedzi NIE, aż w końcu niewielki szum w jego systemie zostaje błędnie wzmocniony do TAK, niosąc katastrofalne rezultaty.

Wiele badań zgrabnie wpasowuje się w ten obraz halucynacji. Jest na przykład dobrze znanym faktem to, że halucynacje są normalnym rezultatem przedłużającej się deprywacji sensorycznej (np. Vosberg, Fraser i Guehl 1960). Możliwe wytłumaczenie jest takie, że w deprywacji sensorycznej część sterowana danymi w systemie generowania i testowania hipotez nie otrzymuje żadnych danych, przez co obniża próg szumu, który zostaje wzmocniony i tworzy przypadkowe schematy sygnałów potwierdzania i obalania, wywołując w końcu szczegółowe halucynacje, których treść jest wytworem tylko i wyłącznie niecierpliwego oczekiwania i przypadkowego potwierdzania. Co więcej, w większości raportów halucynacje rozwijają się stopniowo (pod wpływem deprywacji sensorycznej lub narkotyków). Zaczynają się słabo – na przykład w postaci geometrycznej – a potem stają się silniejsze („przedmiotowe" lub „narracyjne"), a właśnie to ten model przewiduje (np. Siegel i West 1975).

Wreszcie sam fakt, że narkotyk rozprzestrzeniający się w układzie nerwowym może wytworzyć tak złożone i pełne treści efekty, wymaga wyjaśnienia – sam narkotyk z pewnością nie „zawiera opowieści", nawet jeśli naiwniacy chcieliby w to wierzyć. Mało prawdopodobne jest też, że rozprzestrzeniający się w organizmie narkotyk może wytworzyć czy nawet włączyć wyrafinowany system iluzjonistyczny, ale łatwo zauważyć, że narkotyk mógłby zadziałać bezpośrednio, podwyższając, obniżając lub w dowolny sposób zaburzając próg potwierdzania w systemie generowania hipotez.

Model generowania halucynacji zainspirowany grą towarzyską mógłby oczywiście również wyjaśnić strukturę snów. Od czasów Freuda nie ma już wątpliwości co do tego, że tematyczna zawartość snów jest wiele mówiącym symptomem najgłębszych motywacji, lęków i zmartwień śniącego, lecz zawarte we śnie poszlaki są, jak wiadomo, ukryte pod warstwami symboli i błędnych wskazówek. Jakiego rodzaju proces mógłby tworzyć opowieści, które tak skutecznie i nieprzerwanie ujawniają najgłębsze obawy śniącego, jednocześnie ubierając je w warstwy metafor i przemieszczeń? Mniej więcej standardowa odpowiedź freudystów to ekstrawagancka hipoteza wewnętrznego scenarzysty snów, który tworzy terapeutyczne scenariusze dla pożytku ego i w przebiegły sposób wciska je za plecami wewnętrznego cenzora, kamuflując ich prawdziwe znaczenie. (Model freudowski można by nazwać „modelem Hamleta", gdyż przypomina przebiegłą sztuczkę, w której Hamlet wystawił *Pułapkę na myszy* tylko dla

Klaudiusza; trzeba być naprawdę przebiegłym, żeby wymyślić tak subtelny fortel, ale jeśli wierzyć Freudowi, wszyscy nosimy w sobie takiego wirtuoza narracji). Jak zobaczymy później, teorie, które postulują istnienie takiego *homunkulusa* („małego człowieczka" w mózgu), nie zawsze powinny być ignorowane, ale jeżeli homunkulus jest wzywany na pomoc, to lepiej, jeśli okazuje się raczej bezmyślnym funkcjonariuszem niż genialnym freudowskim scenarzystą, który przypuszczalnie produkuje dla nas nowe senne sceny każdej nocy! Analizowany model eliminuje scenarzystę i opiera się na „publiczności" (analogicznej do naiwniaka z gry), która dostarcza treści. Publiczność oczywiście nie jest głupia, ale przynajmniej nie musi mieć teorii swoich własnych lęków; wystarczy, że te lęki prowadzą do kolejnych zadawanych pytań.

Przy okazji warto zauważyć, że jedyna zasada z naszej gry towarzyskiej, która nie byłaby potrzebna w procesie tworzenia snów czy halucynacji, to zasada niesprzeczności. Jako że system percepcyjny przypuszczalnie zawsze eksploruje bieżącą sytuację (która nie jest faktem dokonanym, skończoną senną narracją), kolejne „sprzeczne" potwierdzenia mogą być interpretowane przez maszynerię jako wskazówka zmiany w świecie, a nie jako korekta historii znanej temu, kto sen relacjonuje. Duch był przed chwilą niebieski, a nagle stał się zielony; jego ręce zamieniły się w szpony i tak dalej. Gotowość do metamorfozy, jaką mają przedmioty w snach i halucynacjach, jest jedną z najbardziej uderzających cech tych narracji, a jeszcze dziwniejsze jest to, jak rzadko owymi metamorfozami się we śnie przejmujemy. Gospodarstwo w Vermont okazuje się bankiem w Portoryko, a koń, na którym jechałem, jest teraz samochodem, nie, motorówką, zaś mój kompan zaczynał jazdę jako moja babcia, a teraz jest papieżem. Takie rzeczy się zdarzają.

Owa gotowość jest dokładnie tym, czego spodziewalibyśmy się po czynnym, lecz niewystarczająco sceptycznym pytającym skonfrontowanym z przypadkową próbką odpowiedzi „tak" i „nie". Jednocześnie uporczywość pewnych motywów i przedmiotów w snach, to, że nie poddają się metamorfozom i nie chcą znikać, również może być starannie wyjaśniona przez nasz model. Udając na chwilę, że mózg korzysta z zasady alfabetycznej oraz że pracuje w języku polskim, możemy sobie wyobrazić, jak podświadome pytania prowadzą do obsesyjnego snu:

P: Czy sen jest o ojcu?
O: Nie.
P: Czy jest o telefonie?
O: Tak.
P: Okej. Czy jest o synu?
O: Nie.
P: Czy jest o ojcu?
O: Nie.
P: Czy jest o ojcu rozmawiającym przez telefon?
O: Tak.
P: *Wiedziałem*, że jest o ojcu. Czy ojciec rozmawiał ze mną?
O: Tak. ...

Ten skromny szkic teorii prawdopodobnie nie udowadnia niczego (jeszcze) o halucynacjach czy snach. Pokazuje jednak – metaforycznie – jak *mogłoby* wyglądać mechanistyczne wyjaśnienie tych zjawisk, a to ważny początek, gdyż są osoby skłaniające się ku defetystycznej tezie, że nauka „z założenia" nie może wyjaśnić różnych „tajemnic" umysłu. Przedstawiony szkic nie podejmuje jednak kwestii naszej *świadomości* snów i halucynacji. Co więcej, mimo że pozbyliśmy się mało prawdopodobnego homunkulusa, sprytnego iluzjonisty/scenarzysty, który płata figle w mózgu, na jego miejscu zostawiliśmy nie tylko niemądrych pytaczy-odpowiadaczy (którzy prawdopodobnie mogą być „zastąpieni przez

maszyny"), ale również wciąż dość mądrego pytającego, czyli „publiczność". Może wyeliminowaliśmy sprawcę, lecz nawet nie daliśmy dojść do głosu ofierze.

Poczyniliśmy jednak pewne postępy. Widzimy, jak dbałość o wymagania „inżynieryjne" w zjawisku umysłowym może prowadzić do nowych pytań, na które można z łatwością odpowiedzieć, na przykład: Jakie modele halucynacji mogą uniknąć eksplozji kombinatorycznej? Jak treść takiego przeżycia może powstawać w wyniku (stosunkowo) bezmyślnego, niepojmującego niczego procesu? Jaki rodzaj połączeń między procesami lub systemami mógłby wyjaśnić rezultaty ich interakcji? Jeżeli chcemy stworzyć naukową teorię świadomości, będziemy musieli stawić czoło wielu pytaniom tego typu.

Wprowadziliśmy również zasadniczy zarys tego, co przed nami. Najważniejszym elementem w próbach wyjaśnienia tego, jak w ogóle możliwe są halucynacje i sny, było to, że mózg musi jedynie *zaspokoić epistemologiczny głód* – zaspokoić „ciekawość" we wszelkiej jej postaci. Jeśli ofiara jest bierna lub niezainteresowana tematem *x*, jeśli nie poszukuje odpowiedzi na pytania związane z tematem *x*, wtedy żaden materiał związany z tematem *x* nie musi być przygotowywany. (Jeśli nie swędzi, to nie drap). Świat zapewnia niewyczerpany zalew informacji, które bombardują nasze zmysły, a kiedy skupimy się na tym, ile ich do nas dociera i ile jest nieprzerwanie dla nas dostępnych, często poddajemy się złudzeniu, że wszystkie są cały czas wykorzystywane. Nasze możliwości korzystania z informacji i nasze apetyty epistemologiczne są jednak ograniczone. Jeśli nasze mózgi są w stanie zaspokoić wszystkie nasze poszczególne potrzeby epistemologiczne w momencie, gdy się pojawiają, to nigdy nie będziemy mogli narzekać. Nigdy jednak nie będziemy mogli stwierdzić, czy nasze mózgi dostarczają nam mniej niż wszystko, co jest dostępne w świecie.

Ta zasada prostoty została na razie naszkicowana, lecz nie udowodniłem, że jest prawdziwa. Jak zobaczymy, mózg nie zawsze zresztą korzysta z tej możliwości, ale nie możemy jej przeoczyć. Nie docenia się siły tej zasady w rozwiązywaniu prastarych problemów.

4. I co dalej?

W kolejnych rozdziałach podejmę próbę wyjaśnienia świadomości. Ściślej mówiąc, wyjaśnię rozmaite zjawiska, które tworzą to, co nazywamy świadomością, pokazując, że są one fizycznymi efektami działania mózgu. Pokażę również, jak te działania wyewoluowały oraz w jaki sposób stały się podstawą do pojawienia się złudzenia co do ich własnych mocy i właściwości. Bardzo trudno wyobrazić sobie, że twój umysł to twój mózg – ale nie jest to niemożliwe. Aby móc to pojąć, trzeba poznać naprawdę sporo naukowych odkryć dotyczących działania mózgu, lecz co ważniejsze, trzeba nauczyć się nowych sposobów myślenia. Poznawanie nowych faktów pozwala nam wyobrazić sobie nowe możliwości, ale odkrycia i teorie neuronauki nie wystarczą – nawet neuronaukowcy są często zdumieni świadomością. By pomóc w ćwiczeniu wyobraźni, oprócz naukowych faktów przedstawię też wiele historii, analogii, eksperymentów myślowych oraz innych narzędzi, które dadzą czytelniczkom i czytelnikom nowe perspektywy, zniszczą stare nawyki myślowe i pomogą im uporządkować fakty w jedną spójną wizję znacznie różniącą się od tradycyjnego postrzegania świadomości, do jakiego jesteśmy przyzwyczajeni. Eksperyment myślowy z mózgiem w naczyniu i analogia do gry w psychoanalizę były rozgrzewką przed głównym zadaniem, które będzie polegać na naszkicowaniu teorii biologicznych mechanizmów i *sposobu myślenia* o tych mechanizmach, co pozwoli nam *zobaczyć*, jak można rozwiązać tradycyjne paradoksy i tajemnice świadomości.

W części I omówimy kwestie związane ze świadomością i przyjmiemy pewne metody. Jest to ważniejsze i trudniejsze, niż mogłoby się wydawać. Wiele problemów innych teorii bierze

się z błędnych początków i przedwczesnej chęci znalezienia odpowiedzi na Wielkie Pytania. Nowatorskie założenia mojej teorii odgrywają dużą rolę w tym, co przed nami, a to pozwala nam odłożyć na później wiele tradycyjnych filozoficznych zagadek, o które potykały się inne koncepcje, do czasu zarysowania empirycznej teorii, która zostanie zaprezentowana w części II.

Model wielokrotnych szkiców dotyczący świadomości zarysowany w części II zastępuje tradycyjny model, który nazwałem „teatrem kartezjańskim". Wymaga on radykalnego przemyślenia znanej idei „strumienia świadomości" i początkowo jest głęboko nieintuicyjny, ale z czasem można się do niego przekonać, zwłaszcza gdy zobaczymy, jak wyjaśnia fakty dotyczące mózgu, które dotychczas były ignorowane przez filozofów – i naukowców. Szczegółowo rozważając, jak świadomość mogła wyewoluować, zaczynamy rozumieć wcześniej zdumiewające właściwości naszych umysłów. W części II znajdziemy również analizę roli języka w ludzkiej świadomości oraz związki modelu wielokrotnych szkiców z pewnymi bardziej znanymi koncepcjami umysłu i innymi pracami teoretycznymi z multidyscyplinarnego pola kognitywistyki. Przez cały czas będziemy musieli opierać się kuszącym uproszczeniom charakterystycznym dla tradycyjnego podejścia, zanim postawimy nowe fundamenty.

W części III, mając nowe narzędzia do ćwiczenia wyobraźni, będziemy mogli skonfrontować się (w końcu) z tradycyjnymi tajemnicami świadomości: z dziwnymi właściwościami „pola fenomenologicznego", z naturą introspekcji, z jakością (*qualiami*) stanów przeżywanych, z naturą osobowości czy ego oraz z ich związkami z myślami i wrażeniami, ze świadomością istot innych niż ludzie. Paradoksy związane z tymi kwestiami, które tkwią w tradycyjnych debatach filozoficznych, dostrzeżemy jako powstałe z *niedostatku wyobraźni*, a nie z „rozumienia" – i będziemy w stanie te zagadki rozwiązać.

Książka ta prezentuje teorię, która jest zarówno empiryczna, jak i filozoficzna, a jako że wymogi dla takiej teorii są różnorakie, na końcu książki znajdują się dwa aneksy, które pokrótce mierzą się z wyzwaniami technicznymi, pojawiającymi się w perspektywie naukowej oraz filozoficznej. W następnym rozdziale pytamy, jakie może być wyjaśnienie świadomości oraz czy w ogóle powinniśmy chcieć rozwiązywać zagadki świadomości.

Część pierwsza

Problemy i metody

Rozdział 2

Wyjaśnić świadomość

1. Puszka Pandory: Czy należy zdemistyfikować świadomość?

Oto drzewa i znam ich chropowatość, oto woda i czuję jej smak. Zapach traw, gwiazdy, noc, pewne wieczory, kiedy serce się odpręża – jakże negować ten świat, którego potęgi i siły doznaję? A jednak cała wiedza ziemi nie może mi dać nic, co upewniałoby mnie, że ten świat należy do mnie. Opisujecie mi go i uczycie, jak go klasyfikować. Wyliczacie jego prawa i w moim pragnieniu wiedzy zgadzam się, że są prawdziwe. Rozkładacie jego mechanizm i moja nadzieja rośnie. [...] Na cóż mi tyle wysiłków? Łagodne linie tych wzgórz i ręka wieczoru na niespokojnym sercu nauczą mnie znacznie więcej.

Albert Camus, *Mit Syzyfa*, 1942
[przeł. Joanna Guze]

Przyroda słodycz nauk da ci;
My, wścibską swą mądrością,
Krzywimy piękno wszech postaci:
Mordujemy, aby rozciąć.

William Wordsworth, *Pięknym za nadobne*, 1798
[przeł. Stanisław Kryński]

Ludzka świadomość jest niemalże ostatnią z ocalałych tajemnic. Tajemnica to zjawisko wymykające się ludzkiemu pojmowaniu – jeszcze. Znamy wiele innych tajemnic: tajemnica powstania wszechświata, tajemnica życia i rozmnażania, tajemnica projektu istniejącego w naturze, tajemnica czasu, przestrzeni i grawitacji. Wszystkie one były nie tylko dziedziną naukowej niewiedzy, ale również zadziwiały i zachwycały. Nie mamy jeszcze żadnych ostatecznych odpowiedzi na pytania związane z kosmologią i fizyką cząstek elementarnych, genetyką molekularną czy ewolucjonizmem, ale wiemy już, jak je pojmować. Tajemnice nie zniknęły, lecz zostały okiełznane. Nie przytłaczają już naszego pojmowania, ponieważ wiemy, jak odróżnić pytania niewłaściwe od poprawnych, i nawet jeśli okaże się, że całkowicie myliliśmy się co do obecnie akceptowanych rozwiązań, wiemy już, jak się poruszać w poszukiwaniu lepszych odpowiedzi.

Natomiast w kwestii świadomości nadal panuje straszny zamęt. Świadomość wyróżnia się tym, że jest tematem wprawiającym zwykle nawet najbardziej wyrafinowanych myślicieli w oniemienie i zakłopotanie. A tak jak w przypadku wcześniejszych tajemnic, wielu twierdzi – i to z nadzieją – że nigdy nie nastąpi demistyfikacja świadomości.

Tajemnice po prostu fascynują i dają smak życiu. Nikt nie lubi, gdy ktoś ujawnia sprawcę, zanim inni obejrzą film. Kiedy szydło wyjdzie z worka, nigdy więcej nie odzyskamy już tego cudownego stanu zdumienia, które kiedyś tak nas ogarniało. Niech więc czytelnik uważa. Jeżeli uda mi się osiągnąć cel i wyjaśnić świadomość, to ci, którzy będą czytać dalej, zamienią tajemnicę na fundamenty naukowej wiedzy o świadomości, co nie dla wszystkich może być uczciwą wymianą. Są tacy, dla których demistyfikacja to desakralizacja, i przypuszczam, że od

samego początku odbiorą oni tę książkę jako przejaw intelektualnego wandalizmu, zamach na ostatnie sanktuarium rodzaju ludzkiego. Chciałbym, aby zmienili zdanie.

Camus powiada, że nie potrzebuje nauki, ponieważ więcej mogą go nauczyć łagodne linie wieczornych wzgórz, a ja nie zamierzam się z nim spierać – ze względu na pytania, jakie sobie stawia. Nauka nie potrafi odpowiedzieć na wszystkie ważne pytania. Nie potrafi tego również filozofia. Właśnie z tego powodu zjawisko świadomości, które jest kłopotliwe samo w sobie niezależnie od zmartwień Camusa, nie może być chronione przed nauką – czy raczej przed demistyfikującym, filozoficznym dochodzeniem, które właśnie rozpoczynamy. Ludzie, czasem bojąc się tego, że nauka „morduje, aby rozciąć", jak to określił Wordsworth, fascynują się doktrynami filozoficznymi dającymi taką czy inną gwarancję, że do takiej inwazji nie dojdzie. Obawy, którymi się kierują, mają silne podstawy bez względu na to, jak mocne lub słabe są owe doktryny; demistyfikacja świadomości rzeczywiście *mogłaby* się okazać niepowetowaną stratą. Zamierzam dowieść, że tak się faktycznie nie stanie: jeżeli nawet poniesiemy jakieś straty, to staną się one nieistotne w obliczu wzrostu zrozumienia – naukowego i społecznego, teoretycznego i moralnego – które dobra teoria świadomości może zapewnić.

W jaki jednak sposób demistyfikacja świadomości *mogłaby* stać się czymś, czego żałujemy? Mogłaby być porównywalna do utraty dziecięcej niewinności, co zdecydowanie jest stratą, nawet jeśli dobrze zrekompensowaną. Przyjrzyjmy się na przykład temu, co dzieje się z uczuciem miłości, gdy stajemy się bardziej dojrzali. Rozumiemy, jak mężczyzna w czasach rycerskich mógł chcieć poświęcić życie dla honoru księżniczki, z którą nigdy nie rozmawiał – porywało mnie to, gdy miałem jedenaście czy dwanaście lat – ale nie jest to stan umysłu, w który z chęcią wkroczy osoba dorosła. Kiedyś ludzie mówili i myśleli o miłości w sposób dziś dla nas praktycznie niedostępny – może z wyjątkiem dzieci czy osób jakoś umiejących wygłuszyć to, co wiedzą o świecie. Wszyscy uwielbiamy mówić osobom, które kochamy, że je kochamy, i słyszeć od nich, że jesteśmy kochani – ale jako dorośli przestajemy być pewni, czy wiemy, co to znaczy, tak jak to wiedzieliśmy, kiedy byliśmy dziećmi, a miłość była bardzo prosta.

Czy jest nam lepiej, czy gorzej po tej zmianie perspektywy? Zmiana ta nie jest oczywiście taka sama dla wszystkich. Podczas gdy naiwni dorośli cały czas wznoszą gotyckie romanse na szczyty list bestsellerów, my, wyrafinowani czytelnicy, stwierdzamy, że staliśmy się odporni na zamierzone efekty takich książek – rozśmieszają nas one, a nie doprowadzają do płaczu. A jeśli nawet doprowadzają do płaczu – co się czasem zdarza, mimo żebyśmy tego nie chcieli – jest nam wstyd, gdyż odkrywamy, że wciąż jesteśmy podatni na te tanie sztuczki; tylko niechętnie udaje nam się zrozumieć mizerną bohaterkę martwiącą się o to, czy odnalazła „prawdziwą miłość" – jak gdyby była to jakaś specyficzna substancja (złoto emocjonalne w przeciwieństwie do emocjonalnego mosiądzu lub emocjonalnej miedzi). Takiego dorastania doświadcza nie tylko jednostka. Nasza kultura stała się bardziej wyrafinowana – a przynajmniej to wyrafinowanie jest szerzej rozprzestrzenione w kulturze. W rezultacie zmieniło się nasze postrzeganie miłości, a wraz z tą zmianą zaszły przemiany wrażliwości, przez co nie miewamy przeżyć, które porywały, przygnębiały lub pobudzały naszych przodków.

Coś podobnego dzieje się ze świadomością. Mówimy dziś o naszych świadomych decyzjach i nieświadomych nawykach, o świadomych przeżyciach sprawiających nam radość (w przeciwieństwie do – na przykład – bankomatów, które takich przeżyć nie mają) – ale nie jesteśmy do końca pewni, co mamy na myśli, gdy to mówimy. Cały czas istnieją myśliciele uporczywie twierdzący, że świadomość to jakaś autentycznie cenna rzecz (jak miłość, jak złoto), rzecz po prostu „oczywista" i bardzo, bardzo wyjątkowa, ale pojawia się podejrzenie, że świadomość to iluzja. Być może różne zjawiska wiążą się ze sobą, tworząc w nas poczucie jednego tajemniczego fenomenu, lecz nie cechują się większą zasadniczą jednością niż zjawiska

wywołujące poczucie, że miłość jest prosta.

Porównajmy miłość i świadomość z dwoma innymi zjawiskami – chorobami i trzęsieniami ziemi. Nasze postrzeganie chorób i trzęsień ziemi również zasadniczo się skorygowało przez ostatnie kilkaset lat, ale są one zjawiskami w dużej mierze (choć nie całkowicie) niezależnymi od naszego ich postrzegania. Zmiana podejścia do chorób nie sprawiła sama w sobie, że choroby zniknęły lub stały się rzadsze, ale poskutkowała przemianami w medycynie oraz zdrowiu publicznym i radykalnie zmieniła wzorce występowania chorób. Tak samo trzęsienia ziemi pewnego dnia mogą zostać do jakiegoś stopnia opanowane przez ludzi, a przynajmniej być przez nich przewidywane, lecz na istnienie trzęsień ziemi zdecydowanie nie wpływa nasz stosunek do nich. Z miłością jest inaczej. Nie jest już możliwe dla ludzi dojrzałych „zakochać się" w jakiś ze sposobów możliwych w przeszłości – po prostu dlatego, że nie wierzą oni w te sposoby zakochiwania się. Dla mnie na przykład nie jest już możliwe doświadczenie czystego nastoletniego zauroczenia – chyba że „powrócę do bycia nastolatkiem" i zapomnę lub porzucę większość z tego, co wydaje mi się, że wiem. Na szczęście są inne rodzaje miłości, w które mogę wierzyć, ale gdyby ich nie było? Miłość jest jednym ze zjawisk *zależących od ich postrzegania*, że ujmę to na razie w pewnym uproszczeniu. Istnieją inne, na przykład pieniądze. Gdyby wszyscy zapomnieli o pieniądzach, więcej by ich po prostu nie było; pozostałyby sterty zadrukowanych kawałków papieru, wypukłych metalowych krążków, skomputeryzowane zapisy stanów kont, granitowe i marmurowe budynki bankowe – ale nie byłoby pieniędzy: nie byłoby inflacji, deflacji, kursów wymiany ani odsetek – ani *wartości pieniężnej*. Właściwość owych zadrukowanych kawałków papieru, wyjaśniająca – jak nic innego – ich przekazywanie z ręki do ręki w ramach usług i wymian, wyparowałaby.

W spojrzeniu na świadomość rozwijanym w tej książce okaże się, że jest ona, jak miłość i pieniądze, zjawiskiem, które w ogromnym stopniu zależy od związanych z nią pojęć. Chociaż świadomość, tak jak miłość, ma skomplikowane podłoże biologiczne, to tak jak w przypadku pieniądza niektóre z jej najbardziej znaczących cech są związane z kulturą – nie są po prostu wrodzone w fizyczną strukturę jej przejawów. Jeśli zatem mam rację i jeśli uda mi się obalić niektóre z tych pojęć, zagrożę wymarciem wszelkim zjawiskom świadomości, które od nich zależą. Czy jesteśmy na progu poświadomościowego okresu ludzkiej konceptualizacji? Czy nie powinniśmy się bać? Czy w ogóle da się to wyobrazić?

Gdyby pojęcie świadomości miało „roztrzaskać się w drobny, naukowy mak", co stałoby się z naszym poczuciem odpowiedzialności moralnej i wolnej woli? Gdyby świadome przeżycia miały być w jakiś sposób „zredukowane" do czystej materii w ruchu, co stałoby się z naszym uznaniem miłości, bólu, snów i radości? Jeśli świadome istoty ludzkie miałyby być „tylko" ożywionymi przedmiotami, to jak cokolwiek, co im czynimy, mogłoby być dobre lub złe? Oto obawy rodzące opór i rozpraszające uwagę tych, którzy próbują stawić czoło zamiarom wyjaśnienia świadomości.

Jestem pewien, że te obawy są błędne, choć niekoniecznie w sposób oczywisty. Rośnie stawka w nadchodzącej konfrontacji teorii z argumentami. Istnieją silne przesłanki, niezależne od tych obaw, które przemawiają przeciwko tego rodzaju naukowej, materialistycznej teorii proponowanej przeze mnie, i przyznaję, że to mi przypada w obowiązku udowodnienie nie tylko tego, iż owe argumenty są błędne, ale również tego, że szeroko rozpowszechniona akceptacja mojej wizji świadomości, tak czy owak, nie miałaby tak poważnych konsekwencji. (A gdybym odkrył, że mogłaby mieć takie konsekwencje – co zrobiłbym wówczas? Nie napisałbym *tej* książki, ale poza tym po prostu nie wiem).

Spoglądając na ten problem trochę bardziej optymistycznie, przypomnijmy sobie, co działo się u zarania wcześniejszych demistyfikacji. Nie odebraliśmy ich jako umniejszenia ich

wspaniałości; wręcz przeciwnie: znajdujemy jeszcze głębsze piękno i jeszcze bardziej olśniewającą złożoność wszechświata, niż chroniący owych tajemnic kiedykolwiek w nich odnaleźli. „Magia" wcześniejszych wyobrażeń była przede wszystkim przykrywką dla niedoskonałości naszej wyobraźni, nudnym unikiem, ucieleśnionym w pojęciu *deus ex machina*. Porywczy bogowie poruszający się po niebie złotymi rydwanami to niedorzeczne targowisko z tandetą w porównaniu z olśniewającą niezwykłością współczesnej kosmologii, a złożoność struktury DNA sprawia, że *élan vital* jest niewiele bardziej interesująca od strachu Supermena przed kryptonitem. Gdy rozumiemy świadomość – gdy nie ma już więcej tajemnic – staje się ona inna, ale pozostaje piękno i jeszcze więcej miejsca na zachwyt.

2. Tajemnica świadomości

W czym zatem tkwi tajemnica? Cóż bardziej oczywistego czy pewnego niż to, że on i ona są świadomymi podmiotami przeżyć, że lubią postrzegać i czuć, że znoszą ból, że rozważają różne idee oraz że świadomie podejmują decyzje? Wydaje się to niezaprzeczalne, ale czymże jest świadomość sama w sobie? W jaki sposób ożywione ciała fizyczne mogą wytwarzać takie zjawiska w fizycznym świecie? W tym właśnie tkwi tajemnica.

Zagadka świadomości przejawia się na wiele sposobów, a ostatnio uderzyła mnie z nową siłą o poranku, gdy w bujanym fotelu czytałem książkę. Najwyraźniej właśnie podniosłem wzrok znad książki i początkowo wpatrywałem się ślepo w okno, pogrążony w myślach, kiedy nagle piękno otoczenia wyrwało mnie z zadumy. Była wczesna wiosna, zielonozłote światło słoneczne wpadało przez okno, a tysiące gałązek klonu rosnącego na podwórku było wciąż widocznych spod mgiełki zielonych pączków, tworząc elegancki deseń o wspaniałej strukturze. Szyba w oknie jest ze starego szkła i jest na niej ledwo zauważalna rysa, więc gdy bujałem się w fotelu, ta niedoskonałość spowodowała falę skoordynowanych ugięć, które poruszały się tam i z powrotem poprzez deltę gałęzi; regularny ruch, nałożony z niezwykłą wyrazistością na bardziej chaotyczne migotanie gałązek na wietrze.

Wtedy zdałem sobie sprawę, że ten wizualny metronom w gałęziach drzewa porusza się w rytmie *concerto grosso* Vivaldiego, którego słuchałem w tle. Z początku pomyślałem, że z pewnością nieświadomie zsynchronizowałem moje bujanie się z muzyką – tak jak ktoś mógłby nieświadomie poruszać stopą w rytm muzyki – ale fotele bujane mają ograniczony zakres łatwych do utrzymania częstotliwości bujania, więc ta synchronizacja była raczej przypadkowa i tylko delikatnie udoskonalana przez moją nieświadomą preferencję do elegancji i chęci pozostania w rytmie.

Szybko zacząłem sobie przypominać procesy mózgowe mogące wyjaśnić, w jaki sposób nieświadomie dostosowujemy zachowanie, w tym zachowanie naszych oczu i przekierowywanie uwagi, w celu „zsynchronizowania ścieżki dźwiękowej" z „obrazem", ale rozważania te urwały się, kiedy zdałem sobie sprawę z czegoś innego. *To, co robiłem* – wzajemne oddziaływanie przeżyć i myśli, właśnie opisane przeze mnie z mojego uprzywilejowanego, pierwszoosobowego punktu widzenia – było trudniejsze do „zamodelowania" niż nieświadome, zakulisowe procesy, które bez wątpienia zachodziły *we mnie* i były przyczynami tego, co robiłem. Zakulisowy system był w miarę łatwy do rozgryzienia; to procesy zachodzące w centrum uwagi były absolutnie zdumiewające. Moje świadome myślenie, a szczególnie radość, jaką czułem z połączenia światła słonecznego, słonecznych skrzypiec Vivaldiego, poruszających się gałęzi – oraz przyjemność, jaką dawało mi po prostu myślenie o tym – jak to *wszystko* mogłoby być tylko czymś materialnym zachodzącym w moim mózgu? Jak jakakolwiek kombinacja elektrochemicznych zdarzeń mogłaby składać się na to, jak setki gałązek poddały się muzyce? W jaki sposób jakieś

zdarzenie z zakresu przetwarzania informacji w moim mózgu mogłoby *być* delikatnym ciepłem światła słonecznego opadającego na mnie? Oraz, przede wszystkim, jak pewne zdarzenie w moim mózgu mogłoby *być* wyobrażonym przeze mnie wizerunkiem... jakiegoś innego zdarzenia z zakresu przetwarzania informacji w moim mózgu? Zdaje się to niemożliwe.

Tak, wydaje się, jakoby zdarzenia, które *są* moimi świadomymi myślami i przeżyciami, nie mogły być zdarzeniami mózgowymi, lecz istniało *coś innego*, coś bez wątpienia wytworzonego bądź spowodowanego przez zdarzenia mózgowe, ale stojącego ponad nimi, złożonego z czegoś innego, znajdującego się gdzieś indziej. Właściwie czemu nie?

3. Urok substancji umysłowej

Zobaczmy, co się stanie, gdy podążymy tą niewątpliwie kuszącą ścieżką. Najpierw przeprowadźmy mały eksperyment. Sprowadza się on do zamknięcia oczu i wyobrażenia sobie czegoś, a następnie, gdy już uformujesz wizerunek mentalny i dokładnie go zbadasz, odpowiesz na poniższe pytania. Nie czytaj pytań, zanim nie doczytasz do końca instrukcji: gdy zamkniesz oczy, wyobraź sobie, najszczegółowiej jak potrafisz, fioletową krowę.

Już? A zatem:

(1) Czy twoja krowa miała głowę obróconą w lewo, w prawo, czy patrzyła się do przodu?

(2) Czy coś przeżuwała?

(3) Czy było widać jej wymiona?

(4) Czy była raczej w kolorze jasnofioletowym, czy ciemnofioletowym?

Osoby, które trzymały się instrukcji, prawdopodobnie mogą odpowiedzieć na wszystkie pytania bez konieczności wymyślania czegokolwiek. Jeśli wszystkie cztery pytania okazały się dla kogoś zawstydzająco wymagające, to prawdopodobnie w ogóle nie zadał sobie trudu wyobrażenia sobie fioletowej krowy, a jedynie pomyślał ze znudzeniem: „Wyobrażam sobie teraz fioletową krowę" lub „Powiedzmy, że sobie wyobrażam fioletową krowę", lub coś w tym rodzaju.

Wykonajmy teraz drugie ćwiczenie: zamknij oczy i wyobraź sobie, najszczegółowiej jak potrafisz, *żółtą* krowę.

Tym razem będziesz prawdopodobnie w stanie odpowiedzieć bez oporów na trzy pierwsze pytania oraz będziesz mieć nieco do powiedzenia na temat rodzaju żółtego koloru – pastelowego, maślanego czy brązowożółtego – który pokrywał krowę. Ale tym razem odpowiedz na inne pytanie:

(5) Jaka jest różnica pomiędzy wyobrażaniem sobie fioletowej krowy a wyobrażaniem sobie żółtej krowy?

Odpowiedź jest oczywista: pierwsza krowa jest fioletowa, a druga żółta. Mogą być pomiędzy nimi inne różnice, ale ta jest najistotniejsza. Problem w tym, że są to krowy wyobrażone, a nie krowy prawdziwe, namalowane na obrazie lub też kształty krów na ekranie telewizora, ciężko zatem stwierdzić, co mogłoby być fioletowe w pierwszym przypadku, a żółte w drugim. Nic, co kształtem mogłoby przypominać krowę, nie staje się w twoim mózgu (lub w twoim oku) fioletowe w pierwszym przypadku, a żółte w drugim, a nawet gdyby tak było, to niewiele by to pomogło, gdyż wewnątrz twojej czaszki panują ciemności, a poza tym nie ma tam

żadnych oczu, które rozpoznawałyby kolory.

W twoim mózgu zachodzą procesy ściśle powiązane z konkretnymi wyobrażeniami, więc być może w bliskiej przyszłości neuronaukowiec, badając to, co działo się w twoim mózgu w odpowiedzi na moje instrukcje, będzie w stanie je rozszyfrować do tego stopnia, że zweryfikuje twoje odpowiedzi na pytania od 1 do 4:

„Czy głowa krowy była zwrócona w lewo? Przypuszczamy, że tak. Schemat pobudzenia neuronów związanych z głową krowy był zgodny z prezentacją w lewej górnej ćwiartce wizualnej, zaobserwowaliśmy również sygnały sugerujące ruchy oscylacyjne w zakresie jednego herca, które wskazują na przeżuwanie u krowy, ale nie dostrzegliśmy żadnej aktywności w reprezentacji kompleksu wymion, a po skalibrowaniu potencjałów wywołanych z profilami detekcji kolorów u podmiotu przypuszczamy, że kłamie on co do koloru: wyobrażana krowa była niemal z całą pewnością brązowa".

Wyobraźmy sobie, że wszystko to jest możliwe; załóżmy, że naukowe czytanie w myślach dochodzi do skutku. Mimo to tajemnica nie zostaje rozwiązana: co jest brązowe, gdy wyobrażasz sobie brązową krowę? Nie jest to zdarzenie w mózgu, które naukowcy skalibrowali z twoim przeżyciem koloru brązowego. Rodzaj i położenie zaangażowanych neuronów, ich połączenia z innymi częściami mózgu, częstotliwość lub amplituda aktywności, uwolnione neuroprzekaźniki – żadna z tych właściwości nie jest właściwością krowy „w twojej wyobraźni". A ponieważ wyobrażasz sobie krowę (nie kłamiesz – naukowcy już to potwierdzili), wyimaginowana krowa powstała w tamtym momencie. Ta wyobrażona krowa musi być przedstawiona nie za pośrednictwem substancji mózgowej, ale substancji... umysłowej. Czym innym mogłaby być?

Substancja umysłowa musi zatem być tym, z czego „utkane są sny", i najwyraźniej ma pewne niesamowite właściwości. Jedną z nich już po drodze zauważyliśmy, lecz jest wyjątkowo odporna na zdefiniowanie. Tutaj przede wszystkim należy zaznaczyć, że substancja umysłowa zawsze *ma świadka*. Odnotowaliśmy już, że problem ze zdarzeniami w mózgu jest taki, iż bez względu na to, jak szczegółowo „pasują" do wydarzeń w naszym strumieniu świadomości, mają jedną, najwyraźniej zgubną, wadę: *nikt ich tam nie ogląda*. Zdarzenia, które zachodzą w twoim mózgu, tak jak te zachodzące w twoim żołądku lub wątrobie, zwykle ani nie mają żadnego świadka, ani też nie ma żadnego znaczenia dla ich przebiegu to, czy ktoś się im przygląda, czy nie. Z drugiej zaś strony zdarzenia w świadomości są „z definicji" obserwowane; są *doświadczane* przez *osobę doświadczającą*, i ów fakt sprawia, że są tym, czym są: *świadomymi* zdarzeniami. Świadome zdarzenie najwyraźniej nie może powstać samo z siebie; musi być *czyimś* doświadczeniem. Aby zaistniała jakaś myśl, ktoś (jakiś umysł) musi ją pomyśleć, aby pojawił się ból, ktoś musi go poczuć, a żeby „w wyobraźni" powstała fioletowa krowa, ktoś musi ją sobie wyobrazić.

Problem z mózgami najwyraźniej jest taki, że jak w nie spoglądasz, to odkrywasz, że *nikogo tam nie ma*. Żadna część mózgu nie jest myślicielem, który doprowadza do myślenia, czującym, który doprowadza do czucia, a i mózg jako całość niespecjalnie nadaje się do tej nadzwyczajnej roli. To śliski temat. Czy mózgi myślą? Czy oczy widzą? Czy to ludzie myślą swoimi mózgami i widzą swoimi oczami? Czy to jakaś różnica? Czy jest to banalna kwestia „gramatyki", czy też ujawnia nam główne źródło zamieszania? Idea, że *jaźń* (lub osoba, lub nawet dusza) jest czymś innym niż mózg czy ciało, pozostaje głęboko zakorzeniona w tym, jak mówimy, a więc również w tym, jak myślimy.

Mam mózg.

Taka wypowiedź wydaje się zupełnie niekontrowersyjna. Zdaje się jednak, że nie znaczy tylko:

To ciało ma mózg (i serce, dwa płuca itd.).
lub
Ten mózg ma siebie.

Dosyć naturalne jest myślenie o „jaźni i jej mózgu" (Popper i Eccles 1999) jak o dwóch różnych rzeczach o różnych właściwościach, choćby nawet i były od siebie bardzo uzależnione. Jeśli jaźń nie jest tym samym co mózg, wydaje się, że musi być zrobiona z substancji umysłowej. Po łacinie rzecz myśląca to *res cogitans*, co jest terminem rozpowszechnionym przez Kartezjusza, który podał coś, co w jego mniemaniu było niepodważalnym dowodem na to, że on, oczywista rzecz myśląca, nie może być swoim mózgiem. Poniżej jedna z przekonujących wersji tego dowodu:

Następnie, rozpatrując z uwagą, czym jestem, spostrzegłem, iż o ile mogę sobie przedstawić, że nie mam ciała i nie ma żadnego świata ani miejsca, gdzie bym był, nie mogę sobie jednakowoż przedstawić, jakobym nie istniał wcale. Przeciwnie, z tegoż właśnie, iż zamierzałem wątpić o prawdzie innych rzeczy, wynikało bardzo jasno i pewnie, że istnieję; natomiast, gdybym tylko przestał myśleć, choćby nawet wszystka reszta tego, co sobie wyobraziłem, była prawdą, nie miałbym żadnej przyczyny mniemać, iż istnieję. Poznałem stąd, że jestem substancją, której całą istotą, czyli naturą, jest jeno myślenie, i które, aby istnieć, nie potrzebuje żadnego miejsca ani nie zależy od żadnej rzeczy materialnej. [*Rozprawa o metodzie*, 1637, przeł. Tadeusz Boy-Żeleński]

Odkryliśmy zatem dwa rodzaje rzeczy, które mogłyby składać się z substancji umysłowej: fioletową krowę, która nie znajduje się w mózgu, oraz rzecz, która myśli. Istnieje jednak jeszcze kilka innych szczególnych mocy, które można by przypisać substancji umysłowej.

Wyobraźmy sobie, że winiarnia postanowiła zastąpić swoich ludzkich degustatorów wina maszynami. Komputerowy „system ekspercki" kontroli jakości i klasyfikacji wina w pełni wykorzystuje dostępne technologie. Wiemy już wystarczająco dużo o odnośnych zagadnieniach chemicznych, aby utworzyć przetworniki zastępujące kubki smakowe oraz receptory zapachów nabłonka (dostarczające „surowego materiału" – zewnętrznego bodźca – [zmysłom] smaku i zapachu). Nie znamy szczegółów łączenia się i oddziaływania wzajemnego tych bodźców, tworzących nasze przeżycia, ale na tym polu dokonuje się coraz większy postęp. Praca nad widzeniem poszła o wiele dalej. Badania nad postrzeganiem barw pokazują, że ogromnie trudne technicznie byłoby maszynowe naśladowanie ludzkiej specyfiki, subtelności i niezawodności w zakresie oceny kolorów, lecz nie jest ono niemożliwe. Możemy sobie zatem łatwo wyobrazić użycie zaawansowanych wytworów tych przetworników sensorycznych i ich mechanizmów do tworzenia skomplikowanych klasyfikacji, opisów i schematów oceny. Wlewamy próbkę wina do lejka, a po kilku minutach lub godzinach system wyświetla analizę chemiczną wraz z komentarzem: „Pinot ekstrawagancki i aksamitny, choć brak mu wigoru" – lub inne tego rodzaju słowa. Taka maszyna mogłaby dać sobie radę lepiej niż ludzcy degustatorzy wina podczas wszelkich rozsądnych testów precyzji i spójności, które producenci wina byliby w stanie wymyślić, ale z pewnością bez względu na możliwą „czułość" i „rozdzielczość" takiego systemu wydaje się, że nigdy nie miałby przeżyć i przyjemności, które mamy my, gdy degustujemy wino.

Czy to faktycznie tak oczywiste? Wedle rozmaitych ideologii zebranych pod łącznym mianem *funkcjonalizmu*, gdyby skopiować *całą* „strukturę funkcjonalną" systemu poznawczego ludzkiego degustatora wina (łącznie z pamięcią, celami, wrodzonymi awersjami itd.), odtworzyłoby się w ten sposób również *wszystkie* właściwości umysłowe, łącznie z radością, zachwytem, smakowaniem, które sprawiają, że tak wielu z nas *ceni* sobie picie wina. Zasadniczo nie ma znaczenia, mówi funkcjonalista, czy system jest utworzony z cząstek organicznych, czy z krzemu, dopóki *realizuje to samo zadanie*. Sztuczne serca nie muszą być zrobione z tkanki

organicznej, więc sztuczne mózgi też nie – przynajmniej teoretycznie. Jeśli wszystkie funkcje sterujące mózgu ludzkiego degustatora win mogą być odtworzone z użyciem układów krzemowych, radość zostanie *ipso facto* odtworzona razem z nimi.

Pewna odmiana funkcjonalizmu może w końcu zatriumfować (ta książka będzie bronić jednej z jego wersji), ale z pewnością na pierwszy rzut oka wydaje się on skandaliczny. Zdaje się, że żadna maszyna, bez względu na to, jak dokładnie kopiuje procesy mózgowe ludzkiego degustatora win, nie byłaby w stanie zachwycić się winem, sonatą Beethovena czy meczem koszykówki. Aby coś docenić, potrzeba świadomości – czyli czegoś, czego zwykła maszyna nie posiada. Mózg jest jednak pewnego rodzaju maszyną, organem jak serce, płuca czy nerki, których moc można ostatecznie wyjaśnić mechanicznie. Myśl, że to nie mózg się zachwyca, może wydawać się nie do odparcia; jest to odpowiedzialność (lub przywilej) umysłu. Odtworzenie maszynerii mózgu w krzemowej maszynie nie mogłoby w takim razie dostarczyć prawdziwego zachwytu, ale w najlepszym wypadku jego iluzję lub pozór.

Świadomy umysł nie jest więc tylko miejscem, w którym są kolory czy zapachy, nie jest też tylko rzeczą myślącą. To w nim pojawia się również zachwyt. To ostateczny sędzia, rozstrzygający, dlaczego cokolwiek ma jakąkolwiek wartość. Być może wynika to także z faktu, że świadomy umysł ma być źródłem działań intencjonalnych. Jest przecież jasne – ale czy na pewno? – że jeśli robienie czegoś, co jest *wartościowe*, zależy od świadomości, to *wartościowanie* [*mattering*] (radość, zachwyt, cierpienie, troska) powinno również od niej zależeć. Jeśli lunatyk „nieświadomie" krzywdzi, nie ciąży na nim odpowiedzialność, gdyż w istotnym sensie *on* tego nie zrobił; ruchy jego ciała były nierozerwalnie związane z łańcuchem przyczynowym, który doprowadził do krzywdy, ale były one wynikiem *jego* działania nie w większym stopniu, niż gdyby zrobił komuś krzywdę, wypadając z łóżka. Samo uczestnictwo ciała nie sprawia, że działanie jest intencjonalne; nie sprawia tego też uczestnictwo *sterowane przez struktury mózgowe*, ponieważ ciało lunatyka jest pod oczywistą kontrolą struktur mózgowych lunatyka. Dodać do tego należy świadomość, specjalny składnik zmieniający *zdarzenia* w *działania*[5].

Nie jest winą Wezuwiusza, że zabił twoich najbliższych, a uraza (Strawson 1962) i pogarda nie są rozsądnymi reakcjami – chyba że w jakiś sposób sam siebie przekonasz, iż Wezuwiusz, w przeciwieństwie do powszechnej opinii, jest świadomym podmiotem. Osobliwa rzecz, że w rozpaczy pocieszenie niesie wejście w taki stan umysłu, gniewanie się na „furię" huraganu, wyklinanie raka tak niesprawiedliwie atakującego dziecko czy też przeklinanie „bogów". Dawniej, gdy mowa była o czymś „ożywionym" w przeciwieństwie do „nieożywionego", znaczyło to, że to coś ma duszę (po łacinie *anima*)[6]. Zapewne ogromnym pocieszeniem jest pojmowanie tego, co nas dotyka, jako czegoś ożywionego; to przypuszczalnie dogłębnie biologiczny podstęp, skrót pomagający naszym mózgom, żyjącym pod presją czasu, zorganizować się tak, aby myśleć o rzeczach w danym momencie koniecznych do przetrwania.

Mamy wrodzoną skłonność do traktowania każdej zmieniającej się rzeczy tak, jakby miała duszę (Stafford 1983; Humphrey 1983b, 1986), jednak mimo zwyczajności takiego podejścia wiemy, że przypisywanie (świadomej) duszy Wezuwiuszowi jest przesadą. Pytanie o wyznaczenie granicy jest irytujące i do niego wrócimy, ale wydaje się, że świadomość jest dla nas czymś, co odróżnia nas od zwykłych maszyn. Zwykłe cielesne „odruchy" są „automatyczne" i mechaniczne; mogą być związane z obwodami w mózgu, lecz nie wymagają żadnej interwencji ze strony świadomego umysłu. Dość naturalne jest traktowanie naszych ciał jak kukiełek, którymi „my" sterujemy „od wewnątrz". Sprawiam, że kukiełka macha do publiczności, ruszając palcem; poruszam palcem... ruszając duszą? Te problemy rodzące się w takiej koncepcji są powszechnie znane, ale mimo to zdaje się, że jest ona w jakimś sensie słuszna: nie ma

prawdziwego podmiotu działającego, jeśli nie ma świadomego umysłu, który stoi za czynem. Gdy myślimy w ten sposób o umyśle, najwyraźniej odkrywamy jakieś „wewnętrzne ja" czy „prawdziwe ja". To prawdziwe ja to nie mój mózg; to coś, co *jest właścicielem* mojego mózgu („jaźń i *jej* mózg"). Na biurku Harry'ego Trumana w Gabinecie Owalnym w Białym Domu stał słynny napis: „Odpowiedzialność spada na mnie"[7]. Wydaje się jednak, że na żadną część mózgu nie mogłaby *spadać ostateczna odpowiedzialność*, nie mogłaby ona być ostatecznym źródłem moralnej odpowiedzialności, stojącym na początku łańcucha rozkazów.

Podsumowując, znaleźliśmy cztery powody wiary w substancję umysłową. Wydaje się, że świadomy umysł nie może po prostu *być* mózgiem ani żadną jego konkretną częścią, ponieważ nic w mózgu nie może:

(1) być nośnikiem przedstawiającym fioletową krowę;

(2) być rzeczą myślącą, [domyślnym] *ja* w „Myślę, więc jestem";

(3) zachwycać się winem, nienawidzić rasizmu, kochać kogoś, być źródłem *wartościowania*;

(4) działać z moralną odpowiedzialnością.

Akceptowalna teoria ludzkiej świadomości musi wyjaśniać te cztery fakty sugerujące istnienie substancji umysłowej.

4. Dlaczego porzucono dualizm

Idea umysłu jako czegoś odmiennego od mózgu, zrobionego nie ze zwykłej materii, ale z czegoś nadzwyczajnego, to *dualizm*, który zasłużenie cieszy się złą sławą, pomimo przekonujących racji na jego rzecz, o których już pisałem. Od czasów klasycznego ataku Gilberta Ryle'a (1949/1970) na to, co sam nazwał kartezjańskim „dogmatem ducha w maszynie", dualiści są w defensywie[8]. Przeważający pogląd, różnie wyrażany i podpierany różnymi argumentami, to *materializm*: istnieje tylko jeden rodzaj substancji, czyli *materia* – substancja fizyczna w fizyce, chemii i fizjologii – a umysł jest w jakiś sposób tylko zjawiskiem fizycznym. Innymi słowy, umysł to mózg. Według materialistów możemy (zasadniczo!) wyjaśnić każde zjawisko umysłowe, korzystając z tych samych fizycznych zasad, praw i surowców, które wystarczają do wyjaśnienia radioaktywności, dryftu kontynentalnego, fotosyntezy, rozmnażania, odżywiania czy wzrostu. Wyjaśnienie świadomości bez ulegania syrenim śpiewom dualizmu jest jednym z największych wyzwań wobec tej książki. Co w takim razie jest nie tak z dualizmem? Dlaczego jesteśmy mu tak nieprzychylni?

Tradycyjne wątpliwości co do dualizmu były dobrze znane samemu Kartezjuszowi w XVII wieku i można uczciwie powiedzieć, że ani on, ani żaden z następujących po nim dualistów nigdy przekonująco ich nie rozwiał. Jeśli umysł i ciało są dwiema różnymi rzeczami lub substancjami, to mimo to muszą ze sobą współpracować; organy zmysłowe ciała za pośrednictwem mózgu muszą *informować* umysł, muszą coś mu przesłać lub przedstawić mu pewnego rodzaju postrzeżenia, idee albo dane, po czym umysł, po przemyśleniu sprawy, musi odpowiednio *pokierować* działaniem ciała (w tym mową). Stąd pogląd ten jest często nazywany kartezjańskim interakcjonizmem lub dualizmem interakcjonistycznym. Według Kartezjusza centrum interakcyjnym w mózgu jest szyszynka, czyli *epiphysis*. Ukazana jest ona na schemacie Kartezjusza jako znacznie powiększony zaostrzony owal w środku głowy.

Ryc. 2.1

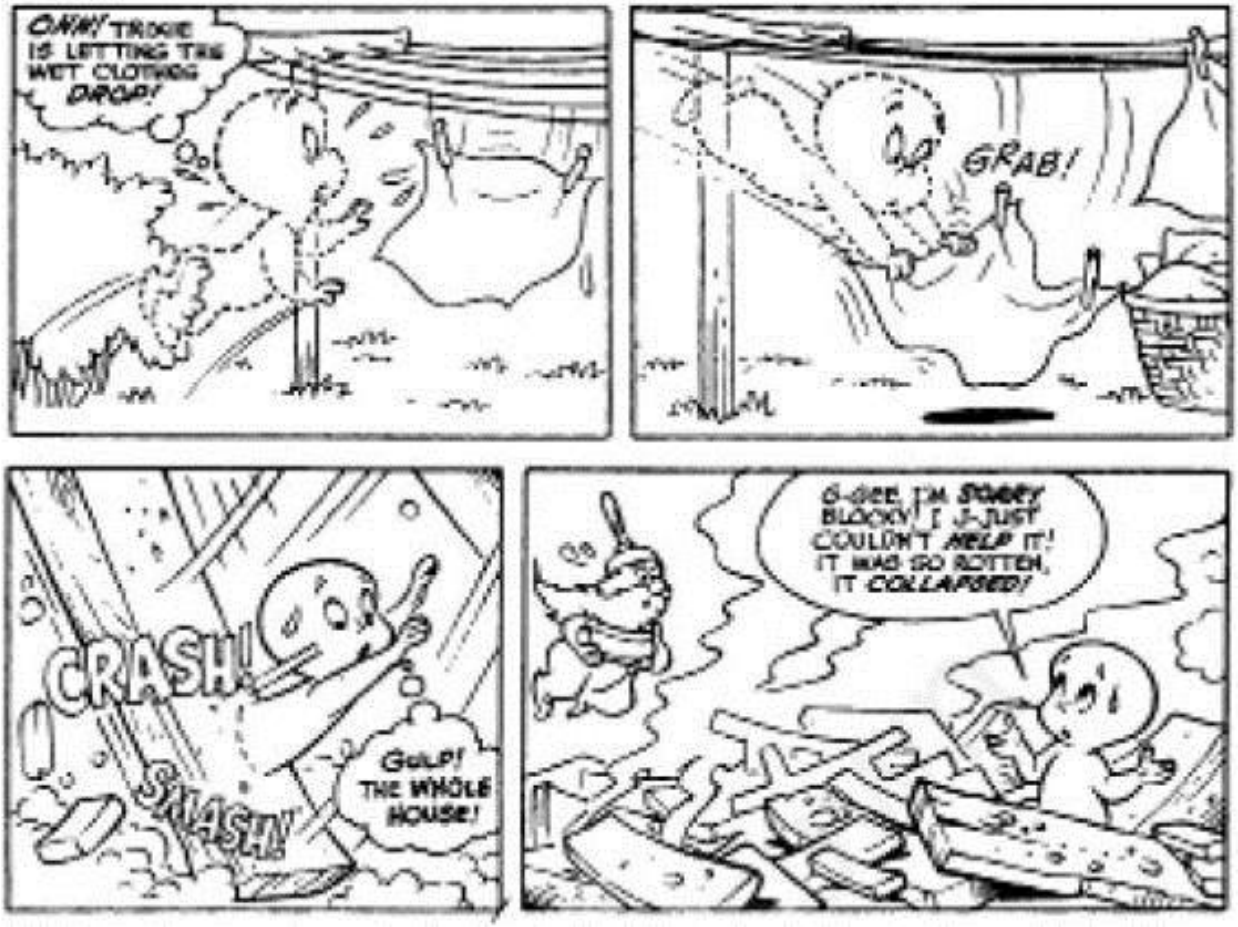
Możemy jasno pokazać problem interakcjonizmu, nakładając rysunek na schemat Kartezjusza.

Ryc. 2.2

Świadoma percepcja strzałki odbywa się dopiero po przekazaniu jakoś przez mózg wiadomości do umysłu, a palec patrzącego może wskazać na strzałkę dopiero po wysłaniu rozkazu do ciała przez umysł. W jaki dokładnie sposób informacja jest przekazywana z szyszynki do umysłu? Jako że nie mamy najmniejszego pojęcia (jeszcze), jakie są właściwości substancji umysłowej, nie możemy nawet zgadywać (jeszcze), jaki wpływ mogłyby na nią wywierać procesy fizyczne wywodzące się z mózgu, więc obecnie je zignorujmy i skoncentrujmy się na sygnałach powracających z umysłu do mózgu. Z założenia nie są one fizyczne; nie są to fale świetlne, fale dźwiękowe, promieniowanie kosmiczne ani strumienie cząstek elementarnych. Nie jest z nimi związana żadna energia ani masa. Jak w takim razie mogą wpłynąć na to, co dzieje się

w komórkach mózgowych, na które muszą oddziaływać, jeśli umysł ma mieć jakiś wpływ na ciało? Fundamentalne zasady fizyki mówią, że jakakolwiek zmiana w torze jakiejkolwiek cząstki fizycznej to przyspieszenie, wymagające nakładu energii. Skąd w takim razie ta energia pochodzi? To właśnie zasadę zachowania energii, która tłumaczy fizyczną niemożliwość istnienia „perpetuum mobile", najwyraźniej narusza dualizm. Od czasu Kartezjusza szeroko omawia się tę konfrontację dość standardowej fizyki z dualizmem. Jest ona zasadniczo postrzegana jako nieuchronna i zgubna wada dualizmu.

Rzecz jasna, omawiano pomysłowe wyjątki techniczne od tej reguły, oparte na wysublimowanej interpretacji fizyki, ale były one mało przekonujące. Upokorzenie dualizmu jest tak naprawdę bardziej oczywiste, niż mogłyby to sugerować wspomniane prawa fizyki. Jest to ta sama niespójność, którą zauważają dzieci – ale tolerują w fikcji – w takich bajkach jak *Kacper i przyjaciele* (Ryc. 2.3). Jak Kacper jest w stanie *zarówno* przepłynąć przez ściany, *jak i* chwycić spadający ręcznik? W jaki sposób substancja umysłowa może *zarówno* uchylać się prawom fizyki, *jak i* sterować ciałem? Duch w maszynie nie jest pomocny w naszych teoriach, chyba że jest to duch, który może przemieszczać rzeczy – niczym głośna zjawa przewracająca lampę lub trzaskająca drzwiami – ale jeśli coś jest w stanie poruszać obiektami fizycznymi, to również jest obiektem fizycznym (choć być może dziwnym i dotychczas niezbadanym rodzajem takiego obiektu).



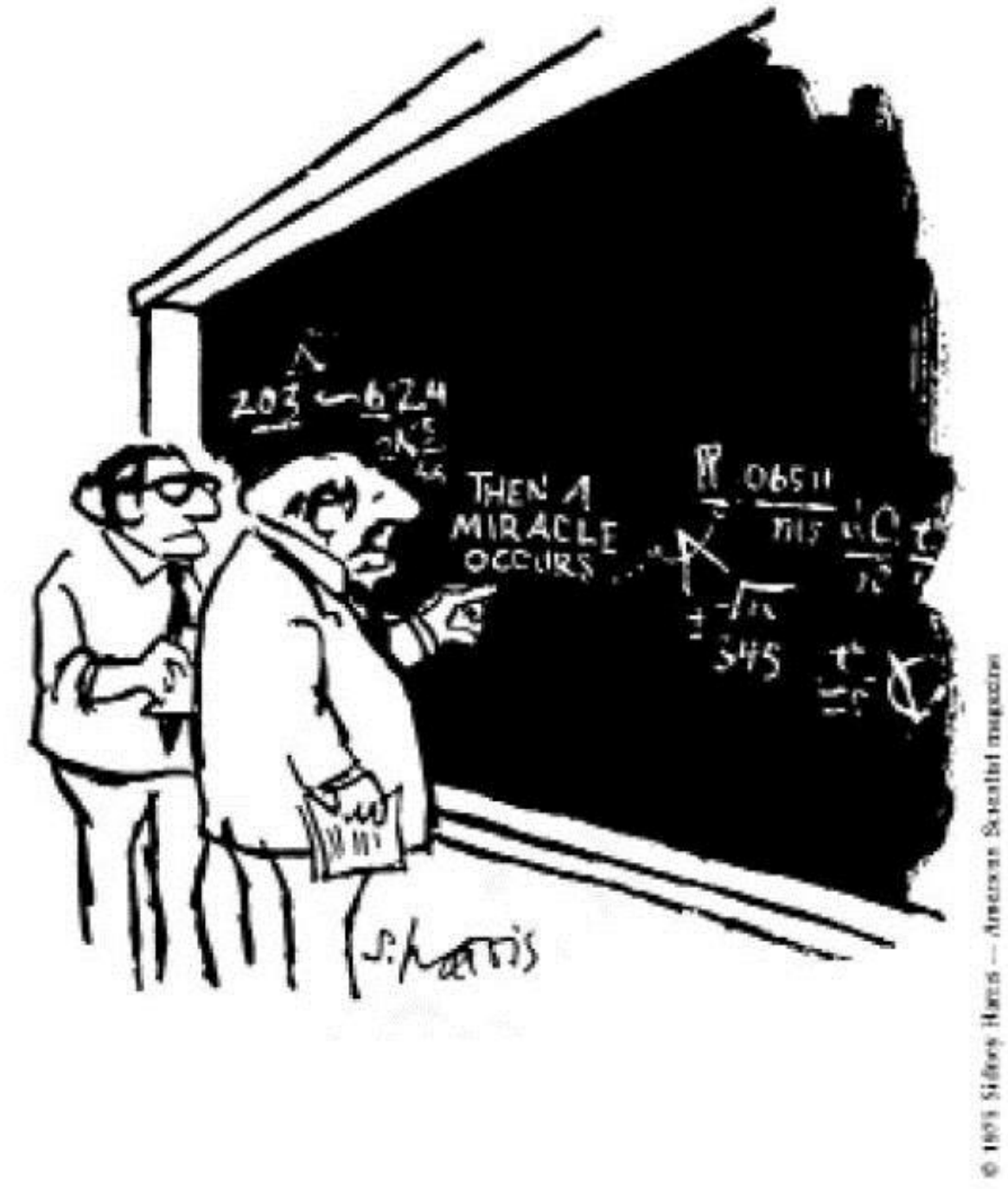
Ryc. 2.3

Cóż w takim razie z możliwością założenia, że substancja umysłowa jest specyficznym rodzajem materii? Podczas wiktoriańskich seansów media często wytwarzały dziwną, lepką substancję zwaną „ektoplazmą", która miała być podstawowym budulcem świata duchowego, a która jednocześnie dawała się zamknąć w szklanym słoiku, sączyła się, wilgotniała i odbijała światło, jak każda znana materia. Te oszukańcze pomysły nie powinny odwodzić nas od postawienia trzeźwego pytania, czy substancja umysłowa mogłaby rzeczywiście być czymś innym niż atomy i cząsteczki tworzące mózg, a jednocześnie materią możliwą do zbadania. Ontologiczne podstawy jakiejś teorii to zbiór przedmiotów i ich rodzajów, które teoria uznaje za istniejące. Ontologia nauk fizycznych uznawała kiedyś „cieplik" (coś, z czego rzekomo składało się ciepło) oraz „eter" (substancja, która przenikała przestrzeń i była nośnikiem fal świetlnych – tak jak powietrze czy woda mogą być nośnikami fal dźwiękowych). Dziś nie są już one traktowane poważnie, natomiast w standardowej ontologii naukowej pojawiły się neutrina, antymateria czy czarne dziury. Być może konieczne jest poszerzenie ontologii nauk fizycznych o pewne pojęcia potrzebne do wyjaśnienia zjawiska świadomości.

Dokładnie taka rewolucja w fizyce została zaproponowana przez fizyka i matematyka Rogera Penrose'a w *Nowym umyśle cesarza* (1989/2008). Choć uważam, że nie udało mu się przekonująco uargumentować rewolucji[9], warto jednak zauważyć, iż autor bardzo uważał, aby *nie* wpaść w pułapkę dualizmu. A co to za różnica? Penrose jasno stwierdza, że proponowana przez niego rewolucja ma zapewnić *łatwiejszy* dostęp do świadomości w badaniach naukowych, nie trudniejszy. Nieprzypadkowo nieliczni dualiści, którzy otwarcie głoszą takie poglądy, szczerze i wygodnie twierdzą, iż nie mają żadnej teorii tłumaczącej pracę umysłu – i obstają przy tym, że jest to poza ludzkimi możliwościami[10]. Pojawia się przypuszczenie, że najatrakcyjniejsza właściwość substancji umysłowej to potencjał pozostania tak tajemniczą, że nauka nigdy nie będzie w stanie jej wyjaśnić.

Ta zdecydowanie antynaukowa postawa dualizmu jest według mnie cechą najbardziej ją dyskwalifikującą oraz powodem, dla którego w tej książce przyjmuję na pozór bezkompromisową zasadę unikania dualizmu *za wszelką cenę*. Nie chodzi o to, że jestem w stanie przedstawić niepodważalny dowód fałszywości i niespójności dualizmu we wszystkich jego postaciach, ale biorąc pod uwagę to, jak upaja się on tajemnicą, stwierdzam, że *akceptacja dualizmu jest kapitulacją* (jak na Ryc. 2.4).

Co do tego stwierdzenia panuje powszechna, choć powierzchowna zgoda, gdyż nie brak problematycznych pęknięć w ścianie materializmu. Naukowcy i filozofowie zgadzają się co do słuszności materializmu, ale – jak zobaczymy – pozbycie się starych wyobrażeń dualistycznych jest trudniejsze, niż by się to współczesnym materialistom mogło wydawać. Znalezienie odpowiednich substytutów tradycyjnych dualistycznych wyobrażeń będzie wymagało zaskakujących zmian w naszym tradycyjnym sposobie myślenia, zmian, które na pierwszy rzut oka mogą się wydawać sprzeczne z logiką zarówno naukowcom, jak i laikom.



Wydaje mi się, że powinieneś sprecyzować krok drugi.

Ryc. 2.3

Moja teoria początkowo zdaje się niezgodna z powszechną wiedzą, ale nie uważam tego za złą wróżbę. Wręcz przeciwnie, od dobrej teorii świadomości nie powinniśmy oczekiwać tego, że będzie miłą lekturą – czymś, co natychmiast jakoś nam się skojarzy, co sprawi, że powiemy sobie z dumą: „Jasne! Cały czas to wiedziałem! To oczywiste, gdy się tak na to spojrzy!”. Gdyby istniała tego rodzaju teoria czekająca na odkrycie, dawno już byśmy na nią wpadli. Tajemnice

umysłu są wokół nas od tak dawna i zrobiliśmy w ich kwestii tak nikły postęp, że jest duże prawdopodobieństwo, iż pewne rzeczy, które zwykliśmy postrzegać jako oczywiste, wcale takimi nie są. Wkrótce powiem, o cóż mi chodzi.

Niektórzy współcześni badacze mózgu – a pewnie ich powściągliwa większość – cały czas udają, że mózg jest dla nich po prostu narządem, takim jak nerka czy trzustka, które powinno być opisywane i wyjaśniane tylko z użyciem najbezpieczniejszych pojęć z zakresu nauk fizycznych i biologicznych. Nie pozwoliliby sobie wspomnieć o umyśle ani o niczym „umysłowym" w swojej pracy. Zdaniem innych badaczy, odważniejszych w teoretyzowaniu, istnieje nowy przedmiot badań: umysł-mózg (Churchland 1986). To nowe pojęcie dobrze oddaje przeważający materializm tych badaczy, którzy chętnie przyznają światu i sobie, że mózg dlatego jest tak interesujący i wyjątkowy, że w jakiś sposób *jest* on umysłem. Jednak nawet wśród tych badaczy panuje niechęć do zmierzenia się z wielkimi problemami, pragnienie odłożenia na później wstydliwych pytań o naturę świadomości.

Tego rodzaju podejście jest dość rozsądne, jest skromnym uznaniem wartości strategii „dziel i rządź", a jednak jego efektem jest zniekształcanie pewnych nowych pojęć powstających na polu tego, co dziś nazywamy *kognitywistyką*. Prawie wszyscy kognitywiści, zarówno ci uważający się za neuronaukowców, jak i psychologowie czy badacze sztucznej inteligencji, zwykle odraczają pytania dotyczące świadomości, ograniczając swoją uwagę do „pobocznych" czy „peryferyjnych" systemów umysłu-mózgu. Te systemy mają zaś dostarczać danych jakiemuś nieokreślonemu „centrum", gdzie następuje „świadome myślenie" i „przeżywanie". To sprawia, że zbyt wiele pracy umysłu pozostawia się owemu „centrum", a to z kolei prowadzi teoretyków do niedoceniania „ilości zrozumienia", jakie musiałyby zapewnić stosunkowo peryferyjne systemy w mózgu (Dennett 1984b).

Wśród naukowców częste jest na przykład przekonanie, że systemy percepcyjne dostarczają „informacji wejściowych" do pewnego rodzaju centralnego pola myślowego, które to następnie „kontroluje" lub „ukierunkowuje" względnie peryferyjne systemy zarządzające ruchem ciała. To centralne pole miałoby również wykorzystywać materiał przetrzymywany w wielu względnie drugorzędnych systemach pamięciowych. Sam jednak pomysł, iż miałyby istnieć istotne podziały teoretyczne między takimi domniemanymi podsystemami jak „pamięć długotrwała" i „rozumowanie" (czy „planowanie"), jest raczej wytworem strategii „dziel i rządź", a nie czymkolwiek, co można odnaleźć w naturze. Jak wkrótce zobaczymy, skierowanie uwagi wyłącznie na konkretne podsystemy umysłu-mózgu często przyczynia się do swoistej krótkowzroczności wśród teoretyków, uniemożliwiającej im dostrzeżenie, że ich modele cały czas zakładają wygodnie ukryty w ciemnym „centrum" umysłu-mózgu teatr kartezjański, miejsce, gdzie „wszystko się łączy" i pojawia się świadomość. Pomysł ten może się wydawać dobry, a nawet nieunikniony, ale dopóki nie przekonamy się, dlaczego tak nie jest, teatr kartezjański będzie nadal przyciągał teoretyków uwięzionych przez iluzję.

5. Wyzwanie

W poprzedniej części rozdziału odnotowałem, że jeśli dualizm jest najlepszą rzeczą, na jaką nas stać, to nie jesteśmy w stanie zrozumieć ludzkiej świadomości. Niektórzy są przekonani, że – tak czy inaczej – jej nie zrozumiemy. Taki pesymizm w obecnym okresie intensywnego rozwoju naukowego wydaje mi się niedorzeczny, a nawet żałosny, lecz przypuszczalnie mógłby być prawdziwy. Być może świadomości naprawdę nie da się wyjaśnić, ale jak możemy być pewni, jeśli nie spróbujemy? Przypuszczam, że wiele – a zapewne większość – kawałków układanki już dobrze rozumiemy, a z moją małą pomocą trzeba je tylko wstawić na właściwe

miejsce. Ci, którzy będą bronić umysłu przed nauką, powinni życzyć mi szczęścia w tej próbie, gdyż, jeśli mają rację, z pewnością nie odniosę sukcesu, jeżeli natomiast wykonam moją pracę najlepiej, jak się da, moja porażka powinna rzucić światło na to, dlaczego nauka nie ma szans na osiągnięcie tego celu. Zdobędą przynajmniej argument przeciwko nauce, a ja wykonam za nich całą brudną robotę.

Podstawowe zasady mojego przedsięwzięcia są proste:

(1) *Żadna cudowna tkanka nie jest dozwolona*. Postaram się wyjaśnić każdą niejasną właściwość ludzkiej świadomości, nie wykraczając poza współczesną fizykę; w żadnym momencie nie będę się odnosił do niewyjaśnionych i nieznanych sił, substancji czy mocy organicznych. Innymi słowy, zamierzam przekonać się, co można zrobić w ramach konserwatywnych limitów standardowej nauki, traktując nawoływania do rewolucji w materializmie jako ostateczność.

(2) *Żadnego symulowania znieczulicy*. O behawiorystach mówi się, że symulują znieczulicę – udają, że nie doświadczają pewnych rzeczy, o których wiemy, że podzielamy je wszyscy. Jeśli będę chciał zaprzeczyć istnieniu jakiejś kontrowersyjnej właściwości świadomości, to na mnie spadnie ciężar *pokazania*, że owa właściwość jest złudzeniem.

(3) *Żadnego czepiania się szczegółów empirycznych*. Postaram się prawidłowo przedstawić wszystkie fakty naukowe, to, co o nich wiemy dzisiaj, ale można się spierać, które naukowe przedsięwzięcia przetrwają próbę czasu. Gdybym miał ograniczać się tylko i wyłącznie do faktów „podręcznikowych", nie mógłbym korzystać z najciekawszych współczesnych odkryć (jeżeli tym właśnie są). A jeśli historia czegoś nas nauczyła, to *i tak* bezwiednie przekazywałbym wam nieprawdę. Niektóre „odkrycia" dotyczące wzroku, za które David Hubel i Torsten Wiesel otrzymali w 1981 roku zasłużoną Nagrodę Nobla, zostały obecnie podważone, a słynna teoria „retinex" Edwina Landa dotycząca postrzegania barw, którą większość filozofów mózgu i innych niespecjalistów uważała za dowiedziony fakt przez ponad dwadzieścia lat, nie jest w najmniejszym stopniu tak poważana wśród naukowców zajmujących się wzrokiem[11].

Jako filozof jestem zainteresowany ustaleniem *możliwości* (oraz obaleniem twierdzeń o niemożliwości), więc zadowolę się nakreśleniem koncepcji, zamiast konstruować teorię wyczerpującą, potwierdzoną empirycznie. Zarys czy model teorii związanej z tym, jak mózg *mógłby* coś robić, przypuszczalnie może zmienić zagadkę w program badawczy: jeśli ten model jest niewystarczający, to może inna, bardziej realistyczna jego odmiana zadziała. (Zarys wyjaśnienia halucynacji w rozdziale 1 jest tego przykładem). Tego rodzaju szkic jest bezpośrednio i wyraźnie podatny na empiryczne podważenie, jednak ktoś, kto twierdzi, że nie jest on *możliwym* wyjaśnieniem zjawiska, musi pokazać, co należy w nim pominąć lub czego *nie może* on zrobić; jeśli po prostu stwierdzi, że mój model może być nieprawidłowy w wielu szczegółach, przyznam się do tych pomyłek. W kartezjańskim dualizmie, na przykład, błędne jest nie to, że Kartezjusz wybrał szyszynkę – a nie na przykład wzgórze czy ciało migdałowate – na centrum interakcji z umysłem, ale *sam pomysł* takiego centrum interakcji umysłowo-mózgowej. Wskazanie na ten błąd zalicza się oczywiście do zmian wynikających z podawania w wątpliwość, co jest związane z rozwojem nauki, a prowadzi do tego, że różni teoretycy tworzą różne standardy. Będę się starał skrajnie przerysowywać moją teorię, nie tylko po to, aby podkreślić kontrast z tradycyjną filozofią umysłu, ale również po to, by dać empirycznym krytykom jaśniejszy cel do atakowania.

W tym rozdziale zapoznaliśmy się z podstawowymi cechami zagadki świadomości. Owa tajemniczość świadomości jest właśnie jedną z jej głównych cech – być może nawet niezbędną cechą, bez której nie przetrwa. Możliwość ta jest szeroko, acz słabo doceniana, a ostrożność zwykle sprzyja doktrynom, które nawet nie próbują wyjaśnić świadomości, gdyż jest ona dla nas tak bardzo znacząca. Dualizm, czyli twierdzenie, że mózg nie może być rzeczą myślącą, a więc rzecz myśląca nie może być mózgiem, jest bardzo kuszący z wielu powodów, ale musimy oprzeć się pokusie; przyjęcie dualizmu jest po prostu cichą kapitulacją. Przyjęcie materializmu nie wiąże się natomiast ani ze zniknięciem problemów dotyczących świadomości, ani z łatwym ułożeniem układanki przez naukę o mózgu. W jakiś sposób umysł musi być mózgiem, a jeżeli nie uda nam się szczegółowo zobaczyć, jak to możliwe, materializm nie rozwiąże tego problemu, lecz jedynie obieca jego rozwiązanie w nieokreślonej przyszłości. Moim zdaniem tej obietnicy nie można dotrzymać, dopóki nie zdobędziemy się na jeszcze bardziej zdecydowane odrzucenie dziedzictwa Kartezjusza. Jednocześnie cokolwiek więcej są w stanie wyjaśnić teorie materialistyczne, nie rozwiążą problemu świadomości, jeśli będziemy zaniedbywać fakty związane z przeżyciami, które tak intymnie poznajemy „od środka”. W kolejnym rozdziale sporządzimy wstępny wykaz tych faktów.

Rozdział 3

Wizyta w ogrodzie fenomenologicznym

1. Witaj w fenomie

Wyobraź sobie szaleńca, który twierdzi, że nie ma czegoś takiego jak zwierzęta. Moglibyśmy spróbować udowodnić mu, że się myli, zabierając go do zoo i mówiąc: „Patrz! Co to jest, jeśli nie zwierzęta?". Nie oczekiwalibyśmy jego ozdrowienia, ale przynajmniej mielibyśmy satysfakcję, że jasno sobie uświadomiliśmy bijące od niego obłąkanie. Wyobraź sobie jednak, że szaleniec mówi: „Ależ ja bardzo dobrze wiem, że istnieją te rzeczy – lwy, strusie i boa dusiciele – ale dlaczego twierdzisz, że te tak zwane zwierzęta to *zwierzęta*? W rzeczywistości są one tylko *robotami* pokrytymi sierścią – właściwie niektóre pokryte są piórami lub łuskami". To wciąż może być obłęd, ale jest to inny, łatwiejszy do obronienia rodzaj obłędu. Ten szaleniec ma po prostu rewolucyjne pojęcie o ostatecznej naturze zwierząt[12].

Zoologowie są ekspertami w kwestii ostatecznej natury zwierząt, a ogrody zoologiczne służą użytecznemu celowi zapoznania społeczeństwa z tymi zagadnieniami. Gdyby zoologowie odkryli, że ten szaleniec ma rację (przynajmniej w pewnym sensie), znaleźliby w zoo świetne zastosowanie do próby wyjaśnienia tego odkrycia. Mogliby powiedzieć: „Okazuje się, że zwierzęta – wiecie: te typowe obiekty, które wszyscy widzimy w zoo – nie są tym, za co je braliśmy. Okazuje się, że są tak inne, że w ogóle nie powinniśmy nazywać ich zwierzętami. Widzicie więc, że *nie istnieją* żadne zwierzęta w zwykłym znaczeniu tego słowa".

Filozofowie i psychologowie często używają terminu *fenomenologia* na oznaczenie ogólnego pojęcia odnoszącego się do wszystkich rzeczy – można by je nazwać „fauną i florą" – które składają się na nasze świadome przeżycia: myśli, zapachy, swędzenia, bóle, wyimaginowane fioletowe krowy, przeczucia i całą resztę[13]. Pojęcie fenomenologii ma kilka różniących się od siebie rodowodów, o których warto wspomnieć. W XVIII wieku Kant odróżnił „fenomeny", czyli to, jak postrzegamy rzeczy, oraz „noumeny", czyli rzeczy same w sobie, a następnie, wraz z rozwojem nauk przyrodniczych w wieku XIX, zaczęto mianem *fenomenologii* określać czysto opisowe badanie materii każdego przedmiotu, neutralne i preteoretyczne. Na przykład fenomenologia magnetyzmu została zapoczątkowana wcześnie, bo w XVI wieku, przez Williama Gilberta, ale jego wyjaśnienie musiało poczekać na odkrycia związków między magnetyzmem i elektrycznością, których dokonali w XIX wieku Faraday, Maxwell i inni. Nawiązująca do owego podziału między przenikliwą obserwacją a teoretycznym wyjaśnieniem szkoła filozoficzna, czy ruch znany jako fenomenologia, pojawiła się na początku XX wieku wraz z pracami Edmunda Husserla. Jej celem było znalezienie nowych podstaw całej filozofii (a właściwie całej wiedzy), co miało się opierać na specjalnej technice introspekcji, według której świat zewnętrzny oraz wszystkie jego następstwa i założenia miały być „wzięte w nawias" w szczególnej czynności umysłu zwanej *epoché*[14]. Zamierzonym rezultatem był badawczy stan umysłu, w którym fenomenolodzy mieli zapoznać się z czystymi przedmiotami świadomego przeżycia, zwanymi *noematami*, nieskażonymi typowymi zniekształceniami i przemianami teorii i praktyki. Tak jak w przypadku impresjonizmu w sztuce czy psychologii introspekcyjnej Wundta, Titchenera i innych, które próbowały zerwać z interpretacją i odsłonić podstawowe fakty dotyczące świadomości w taki sposób, aby można je było ściśle obserwować,

fenomenologii nie udało się znaleźć jednej ustalonej metody, co do której wszyscy byliby zgodni.

O ile zatem istnieją zoologowie, o tyle fenomenologów nie ma: nie ma bezspornych ekspertów do spraw natury rzeczy znajdujących się w naszych strumieniach świadomości. Możemy natomiast podążyć za dotychczasową praktyką i użyć pojęcia fenomenologii jako ogólnego terminu na oznaczenie pewnych obiektów w świadomych przeżyciach, które to obiekty należy wyjaśnić.

Opublikowałem kiedyś artykuł *On the Absence of Phenomenology* [O nieobecności fenomenologii] (1979), w którym opowiedziałem się za szaleństwem rodzaju drugiego: to, z czego składa się świadomość, tak różni się od tego, co myśleliśmy dotychczas, że nie powinniśmy używać starych pojęć. Była to jednak propozycja tak dla niektórych oburzająca („Jak, do licha, moglibyśmy się mylić co do naszego życia wewnętrznego!"), że próbowali ją odrzucić jako przejaw szaleństwa rodzaju pierwszego („Dennett uważa, że nie ma bólu ani zapachów, ani marzeń!"). Trochę to oczywiście przerysowałem, ale nie mogłem się powstrzymać. Moim problemem był brak podręcznego ogrodu fenomenologicznego – w skrócie: fenomu – do wykorzystania w moich wyjaśnieniach. Chciałem powiedzieć: „Okazuje się, że rzeczy pojawiające się w strumieniu świadomości – wiecie: bóle, zapachy, marzenia, obrazy umysłowe, uderzenia złości i pożądania, zwykli mieszkańcy fenomu – nie są tym, czym myśleliśmy, że są. W rzeczywistości są tak inne, że musimy dla nich znaleźć nowe pojęcia".

Zróbmy sobie zatem krótką wycieczkę po ogrodzie fenomenologicznym, aby móc skonstatować, że wiemy, o czym mówimy (nawet jeśli cały czas nie znamy ostatecznej natury tych rzeczy). Wycieczka, celowo powierzchowna i jedynie wprowadzająca, pozwoli nam zwrócić uwagę na pewne rzeczy oraz zadać pewne pytania, zanim zabierzemy się za teoretyzowanie w dalszej części książki. Wkrótce rzucę wyzwanie naszemu codziennemu myśleniu i nie chciałbym, aby ktoś pomyślał, że nie zdaję sobie sprawy ze wspaniałych rzeczy, które dzieją się w umysłach *innych* ludzi.

Nasz fenom dzieli się na trzy części: 1. *przeżycia świata „zewnętrznego"*, takie jak widoki, dźwięki, zapachy, odczucia śliskości i szorstkości, zimna i ciepła czy ułożenia naszych kończyn; 2. *przeżycia czysto „wewnętrznego" świata*, takie jak obrazy wyobraźni, wewnętrzne widoki i dźwięki podczas marzenia o czymś lub mówienia do siebie, wspomnienia, błyskotliwe pomysły, nagłe przeczucia; oraz 3. *przeżycia emocji lub „afektu"* (aby użyć osobliwego terminu, którym posługują się psycholodzy), od bólów cielesnych, łaskotania, uczucia głodu i pragnienia, przez emocjonalne napady złości, radości, nienawiści, zawstydzenia, pożądania, zachwycenia, aż po najmniej cielesne: dumę, lęk, ubolewanie, ironiczny dystans, żal, trwogę, spokój.

Nie domagam się uznania tego trójpodziału na przeżycia zewnętrzne, wewnętrzne i emocjonalne. Niczym zwierzyniec, w którym miesza się nietoperze z ptakami i delfiny z rybami, ta taksonomia zawdzięcza więcej powierzchownym podobieństwom i wątpliwej tradycji niż głębokiemu pokrewieństwu między tymi zjawiskami, ale od czegoś musimy zacząć, a każda taksonomia daje jakąś podporę i pomaga nie przegapić żadnego gatunku.

2. Nasze przeżywanie świata zewnętrznego

Zacznijmy od naszych najpierwotniejszych zmysłów zewnętrznych, mianowicie od smaku i węchu. Jak większość ludzi wie, nasze kubki smakowe są czułe tylko na smaki słodki, kwaśny, słony i gorzki, a przede wszystkim „smakujemy naszymi nosami", i właśnie dlatego jedzenie traci dla nas smak, gdy jesteśmy przeziębieni. Nabłonek węchowy jest dla węchu tym, czym siatkówka oka jest dla wzroku. Pojedyncze komórki nabłonka węchowego różnią się od siebie i każda z nich jest wrażliwa na inny rodzaj cząsteczek powietrza. Ostatecznie liczy się

kształt cząsteczek. Dostają się one do nosa i niczym mikroskopijne klucze uruchamiają konkretne komórki czuciowe w nabłonku. Cząsteczki często mogą być z łatwością wykryte w zdumiewająco niskim stężeniu kilku na miliard. Węch innych zwierząt jest zdecydowanie lepszy od naszego, nie tylko dlatego, że są one w stanie odróżnić więcej zapachów w słabszych tropach (psy gończe są tu najlepszym przykładem), ale posiadają również lepszą czasową i przestrzenną rozdzielczość zapachów. Możemy być w stanie wyczuć w pokoju zapaszek formaldehydu, ale jeśli nam się to uda, to nie czujemy nitkowatego tropu lub regionu, w którym znajdują się wyczuwalne indywidualne cząsteczki; cały pokój, a przynajmniej cały róg pokoju, będzie się wydawał przepełniony zapachem. Nie ma tu żadnej tajemnicy: cząsteczki dostają się przypadkowo do naszych przegród nosowych i w momencie dotarcia do określonych miejsc w nabłonku przekazują skąpą informację o miejscu pochodzenia, w przeciwieństwie do fotonów, które wpadają optycznie prostymi ścieżkami do źrenicy, lądując na specyficznym miejscu na siatkówce, dzięki czemu zostają geometrycznie nakreślone na zewnętrznym źródle lub ścieżce źródłowej. Jeśli rozdzielczość naszego wzroku byłaby tak słaba, jak rozdzielczość naszego węchu, to gdyby przeleciał nad nami ptak, całe niebo by się nim „zajęło". (Niektóre gatunki mają tak słaby wzrok – ich rozdzielczość i umiejętność rozróżniania jest nie lepsza od opisanej – ale to, czym, jeśli czymkolwiek, jest dla zwierzęcia widzieć obiekty tak marnie, stanowi inną kwestię, do której powrócimy w późniejszym rozdziale).

Nasze zmysły smaku i węchu są ze sobą fenomenologicznie połączone, tak jak zmysł dotyku i propriocepcja, czyli zmysł położenia i ruchu naszych kończyn oraz innych części ciała. „Czujemy" obiekty, dotykając, chwytając i popychając je na różne sposoby, ale świadome doznania, które są tego rezultatem, mimo że naiwnie można by je uznawać za zwykłe „przesunięcie" stymulacji receptorów czuciowych pod skórą, są tak naprawdę wytworem rozbudowanego procesu integrowania informacji z wielu źródeł. Zasłoń sobie oczy i weź do ręki patyk (lub długopis czy ołówek). Dotykaj różnych rzeczy wokół siebie tą różdżką i zauważ, z jaką łatwością możesz stwierdzić, jaką mają fakturę – jak gdyby twój układ nerwowy miał czujniki na czubku tej różdżki. Trzeba się nieźle wysilić, choć i tak da to mierne skutki, aby skoncentrować się na odczuciach wywoływanych przez patyk na koniuszkach palców, wibracji patyka i jego oporze w kontakcie z różnymi powierzchniami. Owo porozumienie pomiędzy patykiem a receptorami czuciowymi pod skórą (związane w większości przypadków z ledwo dającymi się usłyszeć dźwiękami) zapewnia informacje, które mózg scala w świadome rozpoznanie faktury papieru, kartonu, wełny czy szkła, jednak te procesy scalania są zupełnie ukryte dla świadomości. Nie zdajemy sobie sprawy z tego, jak to robimy. Weźmy przykład jeszcze większej pośredniości. Pomyśl, w jaki sposób czujesz śliskość oleju na ulicy pod kołami twojego samochodu w momencie skręcania. Fenomenologiczny punkt styku znajduje się w miejscu, w którym guma styka się z jezdnią, nie w żadnym punkcie twojego unerwionego ciała, ubranego i siedzącego w samochodzie, ani na twoich dłoniach w rękawiczkach, opartych na kierownicy.

Nie odsłaniając oczu, poproś kogoś o podanie ci jakiegoś przedmiotu z porcelany, z plastiku, wypolerowanego drewna i metalu. Wszystkie te materiały są wyjątkowo gładkie i śliskie, a mimo to nie będziesz miał problemu z rozróżnieniem tych gładkości – i to nie dlatego, że masz wyspecjalizowane receptory porcelany i plastiku na koniuszkach palców. Wydaje się, że najważniejszym czynnikiem jest przewodność cieplna, ale nie jest ona niezbędna: możesz sam siebie zaskoczyć łatwością, z jaką czasem można rozróżnić te powierzchnie za pomocą różdżki. Jest to prawdopodobnie zależne od wibracji, którym zostaje poddana różdżka, lub od nieopisywalnych – ale wykrywalnych – różnic w usłyszanych trzaskach i dźwiękach drapania. Jednak *masz wrażenie*, że niektóre z twoich zakończeń nerwowych znajdują się w różdżce, gdyż

czujesz różnice w powierzchniach na jej czubku.

Teraz zajmijmy się słuchem. Fenomenologia słuchu to wszystkie dźwięki, które słyszymy: muzyka, wypowiadane słowa, uderzenia, gwizdy, syreny, świergoty i trzaski. Teoretycy zajmujący się dźwiękiem czują czasem pokusę, aby „pozwolić grać małej orkiestrze w głowie". Jest to błąd i abyśmy na pewno go rozpoznali i potem unikali, opowiem wam bajkę.

Dawno, dawno temu, mniej więcej w połowie XIX wieku, zapalony wynalazca podjął debatę z praktycznym filozofem Philem. Wynalazca ogłosił, że zamierza skonstruować urządzenie, które mogłoby automatycznie „nagrywać", a później „odtwarzać" z realistyczną dokładnością, orkiestrę oraz chór wykonujący IX symfonię Beethovena. „Nonsens – powiedział Phil. – To absolutnie niemożliwe. Mogę sobie łatwo wyobrazić mechaniczne urządzenie, które nagrywa sekwencję uderzeń klawiszy pianina, a potem steruje odtwarzaniem tych sekwencji na przygotowanym do tego instrumencie – można by to na przykład zrobić z użyciem rolki dziurkowanego papieru – ale nie ma mowy o ogromnej różnorodności dźwięków i trybów ich odtwarzania przy wykonywaniu IX symfonii Beethovena! Są tam setki ludzkich głosów o różnych donośnościach i tembrach, tuziny instrumentów smyczkowych, dętych blaszanych i drewnianych, perkusyjnych. Urządzenie, które byłoby w stanie odtworzyć taką różnorodność dźwięków, musiałoby być nieporęcznym monstrum przyćmiewającym najpotężniejsze organy kościelne – i gdyby, jak sugerujesz, grało z realistyczną dokładnością, bez wątpienia musiałoby włączyć do pracy dosłownie brygadę niewolników, którzy zajęliby się partiami wokalnymi, a to, co nazywasz »nagraniem« konkretnego występu, ze wszystkimi jego niuansami, musiałoby składać się z setek fragmentów zapisu nutowego – po jednym dla każdego muzyka – z tysiącami, a nawet milionami znaków notacji".

Argument Phila jest cały czas w dziwny sposób nie do odparcia; niesamowite jest, że wszystkie te dźwięki mogą być wiernie przekształcone poprzez transformację Fouriera w spiralny rowek wyżłobiony w płycie winylowej, magnetycznie reprezentowane na taśmie lub optycznie na ścieżce dźwiękowej do filmu. Co więcej, jest jeszcze bardziej zdumiewające, że zwykła tuba, poruszana elektromagnesem przesuwanym spiralną linią, może tak wiernie oddać dźwięk trąbki, brzdąkanie banjo, ludzką mowę czy dźwięk rozbijającej się na chodniku butelki wina. Phil nie był w stanie wyobrazić sobie czegoś tak wspaniałego, a niedostatek wyobraźni wziął za analizę sytuacji.

„Magia" transformacji Fouriera otwiera nowe możliwości do rozpatrzenia, musimy jednak zauważyć, że sama w sobie nie eliminuje problemu, który zadziwiał Phila; jedynie pozwala odłożyć go na później. Podczas gdy my, wysublimowani, możemy śmiać się z Phila, który nie potrafił zrozumieć, jak można nagrać i odtworzyć pewien wzór sprężania i rozprężania powietrza stymulujący ucho, uśmiech zniknie z naszej twarzy, gdy zadamy sobie kolejne pytanie: „Co dzieje się z sygnałem, gdy ucho już go otrzyma?".

Następny zakodowany ciąg modulowanych sygnałów (choć teraz poniekąd zanalizowany i podzielony na równoległe strumienie, złowieszczo przypominające tysiące znaków notacji Phila) wędruje z ucha do ciemnego wnętrza mózgu. Te strumienie sygnałów są *słyszanymi dźwiękami* w takim samym stopniu jak spiralny rowek na płycie; są sekwencjami impulsów elektrochemicznych wędrujących po aksonach neuronów. Czy nie powinno istnieć jeszcze bardziej centralne miejsce w mózgu, gdzie owe ciągi sygnałów sterują działaniem nadrzędnego teatralnego organu umysłu? Kiedy w końcu te bezdźwięczne sygnały są ostatecznie *tłumaczone* na subiektywnie słyszany dźwięk?

Nie chcemy szukać w mózgu miejsc, które wibrują jak gitarowe struny, tak jak nie chcemy tam szukać miejsc, które stają się fioletowe, gdy wyobrażamy sobie fioletową krowę. Są to oczywiste ślepe uliczki, coś, co Gilbert Ryle (1949/1970) nazwałby „błędem kategorialnym".

Co w takim razie *moglibyśmy* znaleźć w mózgu, co dałoby nam satysfakcję, że dotarliśmy do końca historii przeżycia dźwiękowego?[15] W jaki sposób jakikolwiek zbiór fizycznych właściwości wydarzeń w mózgu mógłby być tym samym co ekscytujące cechy dźwięków, które słyszymy? Lub choćby je objaśniać?

Na pierwszy rzut oka te cechy wydają się nie do zanalizowania – lub są, by posłużyć się ulubionym przymiotnikiem fenomenologów, *niewyrażalne*. Jednak przynajmniej niektóre z tych pozornie niepodzielnych i jednorodnych właściwości mogą ujawnić wyraźną złożoność i stać się możliwe do opisania. Weź gitarę i szarpnij pustą strunę basową E (bez naciskania na żaden próg). Uważnie posłuchaj dźwięku. Czy można opisać jego składowe, czy jest może jednym, pełnym dźwiękiem gitarowym? Wielu stwierdzi, że druga możliwość lepiej opisuje ich fenomenologię. Teraz szarpnij otwartą strunę jeszcze raz i ostrożnie przeciągnij palec w dół po progach, aby stworzyć wysoki „szereg harmoniczny". Nagle słyszysz nowy dźwięk: w jakiś sposób „czystszy" i o oktawę wyższy. Niektórzy twierdzą, że jest to zupełnie nowy dźwięk, inni zaś opisują to przeżycie słowami: „dół oderwał się od nuty" – a pozostała tylko góra. Następnie ponownie szarpnij pustą strunę. Tym razem słyszysz, zaskakująco wyraźnie, składową harmoniczną, która została wyizolowana przy drugim pociągnięciu za strunę. Jednorodność i niewyrażalność pierwszego szarpnięcia struny została zastąpiona przez dwoistość równie łatwą do zrozumienia i opisania, jak dwoistość akordu.

Różnice między tymi przeżyciami są zaskakujące, ale złożoność dźwięku zrozumiana przy trzecim szarpnięciu struny istniała od samego początku (mogła potencjalnie zostać dostrzeżona). Badania pokazują, że jesteśmy w stanie odróżnić dźwięk gitary od dźwięku lutni czy klawikordu tylko poprzez skomplikowane szeregi harmoniczne. Tego rodzaju obserwacje mogą nam pomóc *opisać* różne właściwości przeżycia słuchowego przez analizę komponentów informacyjnych oraz procesów, które je ze sobą łączą, pozwalając nam przewidzieć, a nawet sztucznie sprowokować, poszczególne przeżycia słuchowe, lecz nadal nie odpowiadają na pytanie, do czego te właściwości się *sprowadzają*. Dlaczego szereg harmoniczny gitary brzmi *tak*, a lutni *tak*? Nie odpowiedzieliśmy jeszcze na to pytanie, nawet jeśli udało nam się je złagodzić, pokazując przynajmniej niektóre niewyrażalne cechy prowadzące mimo wszystko do jakiejś analizy czy opisu[16].

Badania dotyczące postrzegania słuchowego pokazują, że istnieją specjalistyczne mechanizmy rozszyfrowywania różnych rodzajów dźwięku, trochę jak wyimaginowane komponenty fikcyjnej maszyny odtwarzającej Phila. W szczególności dźwięki mowy są, jak się zdaje, przetwarzane przez coś, co inżynier nazwałby „mechanizmem wyspecjalizowanym". Fenomenologia percepcji mowy pokazuje, że masowa restrukturyzacja tych dźwięków zachodzi w pewnym mózgowym centrum trochę przypominającym inżynierskie studio nagraniowe, w którym łączy się wiele kanałów nagrania, poprawia się je i pod wieloma względami koryguje, aby stworzyć „nagranie matkę" stereo, z którego kolejne nagrania są kopiowane na różne nośniki.

Słyszymy na przykład mowę w naszym języku ojczystym jako sekwencję różnych słów przedzieloną krótkimi, cichymi przerwami. Oznacza to, iż czujemy jasne granice między słowami, które nie mogą zaznaczać się krawędziami kolorów ani liniami, nie są wyróżniane brzdąknięciami czy pstryknięciami, więc pozostałoby sądzić, że granicami muszą być ciche przerwy różnej długości – coś w rodzaju przerw oddzielających litery i wyrazy w alfabecie Morse'a. Gdy badacze proszą badanego o zaznaczenie przerwy pomiędzy słowami, raczej nikt nie ma problemu z dostosowaniem się do prośby. Wydaje się, że przerwy istnieją. Gdy jednak spojrzymy na profil energii akustycznej w sygnale wejściowym, regiony najniższej energii (chwile najbliższe ciszy) wcale nie pojawiają się na granicy słów. Segmentacja dźwięków mowy to proces, w którym granice stawia się na podstawie gramatycznej struktury języka, a nie

fizycznej struktury fali akustycznej (Liberman i Studdert-Kennedy 1977). Pomaga nam to zrozumieć, dlaczego mowę w obcym języku postrzegamy jako pomieszany, nieposegregowany napływ dźwięków: wyspecjalizowane mechanizmy w „studiu dźwiękowym" mózgu nie mają odpowiednich kompetencji gramatycznych, aby wyodrębnić odpowiednie segmenty, więc co najwyżej mogą przekazać nieretuszowaną wersję docierającego sygnału.

Gdy postrzegamy mowę, jesteśmy świadomi nie tylko znaczenia czy gramatycznej kategorii słów. (Gdyby tak było, nie bylibyśmy w stanie rozróżnić słyszenia od czytania). Słowa jednoznacznie od siebie się oddzielają, są uporządkowane oraz zidentyfikowane, ale mają również cechy zmysłowe. Właśnie przed chwilą usłyszałem na przykład wyrazisty brytyjski głos Nicka Humphreya, delikatnie wyzywający, niekoniecznie kpiący. Wydaje mi się, że *słyszę* jego uśmiech, a w moim przeżyciu jest też poczucie, że pomiędzy słowami jest śmiech, który chce się wyrwać jak słońce spoza chmur. Właściwości, których jesteśmy świadomi, to nie tylko rosnąca i opadająca intonacja, ale też chrypienie, świst i seplenienie, nie wspominając o zgryźliwości, drżeniu ze strachu, bezbarwności, depresji. I tak jak właśnie zaobserwowaliśmy w przypadku gitary, coś, co początkowo wydaje nam się całościową, niepodzielną cechą, często z pomocą drobnego eksperymentu i izolacji może zostać poddane analizie. Bez żadnego wysiłku rozpoznajemy pytajne brzmienie pytania – oraz różnicę pomiędzy brytyjskim i amerykańskim dźwiękiem pytajnym – ale potrzebujemy nieco eksperymentowania z wątkami i odmianami pytań, zanim będziemy mogli pewnie i rzetelnie opisać różnice w formie intonacji, które prowadzą do różnych słuchowych smaczków.

„Smaczki" wydają się tu adekwatną metaforą, ponieważ niewątpliwie nasze możliwości analizy smaków są tak ograniczone. Znane, choć cały czas zaskakujące dowody na to, że smakujemy nosami, pokazują, iż nasze moce smakowe i węchowe są tak prymitywne, że mamy problem nawet ze zidentyfikowaniem drogi, którą dociera do nas informacja. Ta niewiedza nie ogranicza się do smaku i węchu; nasz słuch na bardzo niskich częstotliwościach – takich jak najniższe nuty basowe zagrane na kościelnych organach – jest najwyraźniej bardziej spowodowany czuciem wibracji ciała, a nie wibracji w uchu. Niesamowite jest to, że „dźwięk fis, dokładnie dwie oktawy poniżej fis, który mogę zaśpiewać" może być faktycznie *usłyszany* nie przez moje uszy, ale przez moje ciało.

Na koniec przyjrzyjmy się przez chwilę wzrokowi. Gdy nasze oczy są otwarte, mamy poczucie szerokiego pola – często nazywanego polem fenomenalnym lub polem widzenia – w którym pojawiają się obiekty w różnych kolorach, o różnych głębiach i w różnych odległościach, poruszające się lub pozostające w bezruchu. Naiwnie postrzegamy prawie wszystkie przeżywane cechy jako obiektywne właściwości zewnętrznych obiektów, obserwowanych przez nas „bezpośrednio", ale nawet jako dzieci szybko rozpoznajemy przejściowe kategorie przedmiotów – świecenie, błyski, migotanie, rozmyte krawędzie – o których wiemy, że są w pewnym sensie wynikiem interakcji pomiędzy obiektami, światłem oraz naszym aparatem wzrokowym. Mimo wszystko postrzegamy te obiekty „poza nami", nie w nas, z kilkoma wyjątkami: ból spoglądania na słońce lub na nagłe jasne światło, gdy nasze oczy są przyzwyczajone do ciemności, czy wzbudzający obrzydzenie ruch pola fenomenalnego, gdy kręci się nam w głowie. Te przeżycia opisujemy raczej jako „odczucia w oczach", bardziej przypominające naciski i swędzenie odczuwane, gdy trzemy oczy, niż jako normalne właściwości obiektów, które obserwujemy przed nami.

Jedną z rzeczy, które obserwujemy w zewnętrznym świecie fizycznym, są obrazy. Są one w tak oczywisty sposób wyrazistymi przedmiotami do oglądania, że często zapominamy, iż są nowością w naszym środowisku wizualnym, gdyż powstały jakieś kilkadziesiąt tysięcy lat temu. Dzięki niedawnemu pojawieniu się sztuki i rzemiosła jesteśmy obecnie otoczeni obrazkami,

mapami, wykresami, zarówno statycznymi, jak i ruchomymi. Te wizerunki fizyczne, które są jedynie jednym z wielu rodzajów „surowca" dla procesów percepcji wzrokowej, stały się niemal niemożliwym do odparcia modelem „wytworu końcowego" percepcji wzrokowej: „obrazami w głowie". Chętnie mówimy: „Oczywiście, że rezultatem *widzenia* jest obraz w głowie (lub w umyśle). Cóż innego mogłoby to być? Z pewnością nie melodia ani smak!". Tę kuriozalną i powszechną chorobę wyobraźni będziemy jeszcze na kilka sposobów leczyli, ale zacznijmy od przypomnienia: galerie malarstwa dla niewidomych to strata sił i środków, więc obrazy w głowie wymagałyby oczu w głowie, które mogłyby te obrazy docenić (nie wspominając o dobrym oświetleniu). Załóżmy jednak, że istnieją oczy umysłu, które doceniają obrazy powstające w głowie. Cóż te wewnętrzne oczy tworzą z obrazów wewnątrz głowy? Jak mamy uniknąć nieskończonego regresu obrazów i widzów? Może on być przerwany tylko poprzez odkrycie jakiegoś widza, którego postrzeganie unika tworzenia kolejnego obrazu, który z kolei znów potrzebowałby widza. Być może miejsce, w którym możemy przerwać ten regres, to pierwszy krok?

Na szczęście istnieją niezależne powody do sceptycznego podejścia do postrzegania wzroku jako obrazów w głowie. Gdyby wzrok był związany z obrazami w głowie, z którymi my (lub nasze jaźnie) jesteśmy ściśle zaznajomieni, to czy rysowanie nie powinno być prostsze? Przypomnij sobie, jak trudno jest realistycznie narysować na przykład różę w wazonie. Przed tobą znajduje się wielka róża – załóżmy, że na lewo od twojego notesu. (Bardzo proszę o dokładne wyobrażenie sobie tego!) Wydaje się, że wszystkie widoczne szczegóły róży są żywe i wyraźne oraz ściśle dla ciebie dostępne, ale przypuszczalnie łatwy proces przenoszenia czarno-białej, dwuwymiarowej kopii wszystkich tych szczegółów o kilka stopni w prawo jest dla wielu ludzi takim wyzwaniem, że szybko się poddają i stwierdzają, że po prostu nie potrafią rysować. Zmiana trzech wymiarów na dwa jest dla ludzi szczególnie trudna, co jest dość zadziwiające, gdyż odruchowo i bez wysiłku robimy coś, co wydaje się zmianą przeciwną – widzenie dwuwymiarowego obrazu *jako* trójwymiarowej sytuacji czy przedmiotu. Właśnie ze względu na trudność, którą sprawia nam powstrzymanie przeciwnej zmiany, kopiowanie zwykłego obrazka jest tak wymagającym zadaniem.

Nie jest to tylko kwestia „koordynacji wzrokowo-ruchowej", ponieważ ludzie, którzy potrafią zręcznie i bez wysiłku haftować lub składać zegarki kieszonkowe, mogą i tak beznadziejnie źle kopiować obrazki. Ktoś mógłby powiedzieć, że jest to raczej kwestia koordynacji wzrokowo-mózgowej. Osoby, które opanowały tę sztukę, wiedzą, że wymaga ona specyficznych umiejętności koncentracji, sztuczek takich jak delikatna zmiana ogniskowej w oku, aby w jakiś sposób oddalić od siebie to, co się wie (moneta jest okrągła, blat stołu jest prostokątny), aby można było zaobserwować faktyczne kąty zakreślane przez linie na obrazku (moneta jest eliptyczna, blat jest trapezoidalny). Często pomaga nałożenie wyobrażonej siatki lub nawet tylko dwóch skrzyżowanych linii, aby móc ocenić faktyczny kąt pomiędzy obserwowanymi kreskami. Uczenie się rysowania opiera się głównie na uczeniu się, jak zlekceważyć normalne procesy wzrokowe, aby przeżywanie obiektu było *jak patrzenie na obrazek*. Nigdy nie będzie to tylko jak patrzenie na obrazek, ale gdy już zostało rozpracowane w tym kierunku, można, z użyciem innych sztuczek znanych ekspertom, mniej więcej „skopiować" na papier to, co przeżywamy.

Naiwnie pojmowane pole wzrokowe wydaje się jednolicie szczegółowe i skupione od centrum do swoich granic, jednak prosty eksperyment pokazuje, że tak nie jest. Weź talię kart i usuń na bok zakrytą kartę tak, aby jej nie widzieć. Wyciągnij ją przed siebie, zaczynając od lewego lub prawego krańca pola wzrokowego, i obróć ją figurą do przodu, uważając, aby patrzeć cały czas prosto przed siebie (wybierz jakiś punkt, na który będziesz patrzeć). Okaże się, że nie

jesteś w stanie stwierdzić, czy jest to karta czerwona, czarna czy figura. Zauważ jednak, że zdajesz sobie dobrze sprawę z każdego, nawet drobnego ruchu karty. Widzisz ruch, ale bez możliwości dostrzeżenia kształtu czy koloru poruszającego się obiektu. Teraz zacznij przesuwać kartę w kierunku centrum pola wzrokowego, cały czas uważając, aby patrzeć w to samo miejsce przed siebie. W którym momencie jesteś w stanie zidentyfikować kolor? A kiedy oznaczenie i numer? Zauważ, że jesteś w stanie stwierdzić, iż jest to karta z figurą, dużo wcześniej, nim możesz powiedzieć, czy to walet, dama czy król. Zaskoczy cię pewnie, jak blisko centrum możesz przybliżyć kartę, wciąż nie będąc w stanie jej zidentyfikować.

Nie zauważamy tego szokującego deficytu w naszym widzeniu peryferyjnym (całe widzenie oprócz dwóch czy trzech stopni wokół ścisłego centrum), ponieważ nasze oczy, w przeciwieństwie do kamer telewizyjnych, nie są równomiernie wycelowane na świat, ale poruszają się w koło w nieustającej i w dużej mierze niezauważalnej grze w wizualnego berka z obiektami, którymi potencjalnie możemy się zainteresować, a które znajdują się w naszym polu widzenia. Zarówno poprzez łagodne śledzenie, jak i poprzez skoki zwane *sakkadowymi*, oczy dostarczają mózgom wysokiej rozdzielczości informacje o tym, co w danym momencie zajmuje obszar dołka środkowego siatkówki oka. (Dołek środkowy rozróżnia około dziesięć razy lepiej niż obszary siatkówki wokół niego).

Nasza wizualna fenomenologia, *treść* przeżycia wzrokowego, ma formę inną od każdej innej formy reprezentacji, zarówno od obrazu, filmu, mapy, makiety, jak i wykresu. Zastanów się, co jest obecne w twoim przeżyciu, gdy patrzysz na stadion sportowy zapełniony tysiącami widzów. Osoby są od ciebie zbyt oddalone, aby je rozróżnić, chyba że pomoże ci jakaś wyraźna, duża cecha (prezydent – tak, od razu widać, że to on, we własnej osobie; to ten, którego można łatwo dostrzec na tle czerwono-biało-niebieskiej chorągiewki). Widzisz, że tłum składa się z istot ludzkich, bo się poruszają jak ludzie. W przeżyciu wzrokowym tłumu jest coś globalnego (wszystko tam wygląda bardzo „tłumnie", tak samo jak małe drzewko widziane z okna może wyraźnie wyglądać jak wiąz, a podłoga może wydawać się zakurzona), ale nie widzisz wielkiej kropy z napisem „tłum"; widzisz jednocześnie tysiące szczegółów: podskakiwanie czerwonych czapek i połyskujące okulary słoneczne, fragmenty niebieskiego płaszcza, broszury, którymi ktoś macha w powietrzu, i uniesione pięści. Gdybyśmy spróbowali namalować „impresjonistyczne" przedstawienie naszego przeżycia, pobrzękujący ruch kolorowych plamek *nie* uchwyciłby zawartości; nie masz przeżycia pobrzękującego ruchu kolorowych plamek, tak samo jak nie masz przeżycia elipsy, gdy patrzysz z ukosa na monetę. Malowidła – kolorowe dwuwymiarowe obrazy – mogą być mniej lub bardziej podobne do informacji docierających do siatkówki ze sceny trójwymiarowej, a zatem sprawiać wrażenie podobne do tego, gdy przyglądasz się jakiejś rzeczywistej scenie, ale malowidło nie jest w takim razie odwzorowaniem powstałego wrażenia, lecz raczej czymś, co wywołuje lub stymuluje takie wrażenie.

Nie jesteśmy w stanie przedstawić ujęcia fenomenologii wizualnej bardziej realistycznie niż fenomenologii sprawiedliwości, melodii czy szczęścia. Jednak wydaje się często na miejscu, a nawet bywa wręcz pokusą nie do odparcia, żeby przeżycia wzrokowe opisywać jako obrazy w głowie. Ta pokusa również należy do fenomenologii wizualnej, a zatem musi zostać wyjaśniona w kolejnych rozdziałach.

3. Nasze przeżycia świata wewnętrznego

Cóż jest więc „surowcem" naszego życia wewnętrznego i co z nim robimy? Odpowiedź narzuca się sama; przecież „patrzymy i widzimy", a potem spisujemy rezultaty.

Według wciąż żywej tradycji brytyjskich empirystów, Locke'a, Berkeleya i Hume'a,

zmysły są wrotami umysłu; gdy materiał zmysłowy już jest w środku, można go obrabiać i dowolnie łączyć, aby tworzyć wewnętrzny świat rzeczy wyobrażonych. Fioletową, latającą krowę wyobrażasz sobie w ten sposób, że bierzesz kolor fioletowy, który znasz z winogron, skrzydła widzianego wcześniej orła i łączysz je z krową, widzianą wcześniej jako krowę. To jednak nie może być do końca prawda. Do oka dociera promieniowanie elektromagnetyczne, a tym promieniowaniem nie sposób malować wyimaginowanych krów. Nasze narządy zmysłowe są bombardowane różnymi formami fizycznej energii, która jest w miejscu kontaktu „przetwarzana" w impulsy nerwowe. Te impulsy następnie przemieszczają się do wnętrza mózgu. Z zewnątrz do wewnątrz dostaje się jedynie informacja i nawet jeśli jej odbiór może *spowodować* wytworzenie jakiegoś fenomenologicznego szczegółu (mówiąc tak neutralnie, jak to możliwe), trudno uwierzyć, by informacja sama w sobie – która jest abstrakcją przetworzoną w coś faktycznego w pewnym zmodyfikowanym, fizycznym przekaźniku – mogła *być* owym fenomenologicznym szczegółem. Mimo to istnieje powód, aby zgodzić się z brytyjskimi empirystami, że świat wewnętrzny jest w *pewnym sensie* zależny od źródeł zmysłowych.

Wzrok to modalność sensoryczna, którą my, ludzie myślący, prawie zawsze wyróżniamy jako najważniejsze źródło wiedzy zmysłowej, mimo że chętnie uciekamy się do dotyku i słuchu, by potwierdzić to, co mówią nam oczy. Nawyk postrzegania umysłu przez metaforę widzenia (przyzwyczajenie, któremu poddałem się dwa razy tylko w tym zdaniu) to, jak zobaczymy, ważne źródło kłopotów i zamieszania. Wzrok tak bardzo zdominował nasze praktyki intelektualne, że jest nam niezwykle trudno pojmować w inny sposób. W celu osiągnięcia porozumienia przygotowujemy widoczne wykresy i tabele, tak abyśmy mogli „zobaczyć, co się dzieje", a jeśli chcemy „zobaczyć, czy coś jest możliwe", staramy się wyobrazić to sobie „oczyma wyobraźni". Czy gatunek niewidomych istot myślących, posługujących się głównie słuchem, byłby w stanie pojmować za pośrednictwem melodii, dzwonienia i skrzeczenia w uszach wszystko to, co my pojmujemy dzięki „obrazom" umysłowym?

Nawet niewidomi od urodzenia korzystają z wizualnego słownictwa do opisywania swoich procesów myślowych, choć nie jest jeszcze do końca jasne, w jakim stopniu jest to wynikiem ich dostosowywania się do języka, którego uczą się od osób widzących, a w jakim rozpoznaną przez nich stosownością metafory, pomimo jej rozbieżności z ich procesami myślowymi. Może nawet używają w przybliżeniu tych samych mechanizmów wizualnych w mózgu, z jakich korzystają osoby widzące – mimo że brakuje im zwykłych źródeł informacji wzrokowych. Odpowiedzi na te pytania rzuciłyby cenne światło na naturę normalnej ludzkiej świadomości, gdyż jej cechą charakterystyczną jest w dużej mierze wystrój wizualny.

Gdy ktoś nam coś wyjaśnia, możemy zakomunikować nasze zrozumienie, mówiąc „jasne!"[17], i nie jest to tylko martwa metafora. Quasi-wizualna natura fenomenologii rozumienia jest niemal całkowicie ignorowana przez kognitywistów, zwłaszcza badaczy w dziedzinie sztucznej inteligencji (ang. *Artificial Intelligence* – AI), którzy pracują nad stworzeniem systemów komputerowych rozumiejących język. Dlaczego lekceważą fenomenologię? Prawdopodobnie przez swoje przekonanie, że niezależnie od tego, jak rzeczywista i fascynująca pozostaje, jest ona niefunkcjonalna – jest kołem, które się toczy, ale nie angażuje żadnej ważnej maszynerii zrozumienia.

Fenomenologie słuchaczy, będące reakcją na tę samą wypowiedź, mogą różnić się od siebie niemal nieskończenie, bez widocznej różnicy w zrozumieniu czy pojmowaniu. Łatwo sobie wyobrazić, jak różne mogłyby być obrazy umysłowe wywołane u dwóch różnych osób, które słyszą zdanie:

Wczoraj mój wujek zwolnił swojego prawnika.

Jim mógłby zacząć od żywego przypominania sobie *wczorajszych* wydarzeń,

jednocześnie szybko przywołując stopień pokrewieństwa z wujkiem (brat ojca albo matki, lub mąż siostry ojca bądź matki), a następnie wygląd schodów sądowych oraz rozgniewanego staruszka. Natomiast Sally mogłaby pominąć „wczoraj" i skupić uwagę na pewnej anomalii na twarzy swojego wujka Billa, jednocześnie wyobrażając sobie trzaskanie drzwiami i ledwie „widoczny" odjazd elegancko ubranej kobiety z etykietką „prawnik". Niezależnie od swoich obrazów umysłowych, Jim i Sally zrozumieli zdanie tak samo poprawnie, co mogłoby zostać potwierdzone serią późniejszych parafraz oraz odpowiedzi na pytania. Bardziej teoretycznie ukierunkowani badacze zwróciliby poza tym uwagę na to, że obrazy *nie mogły* być kluczem do zrozumienia, gdyż nie możesz narysować wujka, wczoraj, zwalniania czy prawnika. Wujkowie, w przeciwieństwie do klaunów czy strażaków, nie odróżniają się w żaden charakterystyczny, widoczny sposób od innych, natomiast wczoraj nie jest podobne do czegokolwiek innego. Rozumienie nie może być zatem osiągnięte poprzez sprowadzanie wszystkiego do waluty obrazkowej, chyba że reprezentowane obiekty można jakoś zidentyfikować, jak na przykład przez etykietki, jednak tekst na tych etykietkach musiałby być zrozumiany, co znów cofa nas do początku.

To, czy *usłyszę* to, co mówisz, jest uzależnione od tego, czy powiesz to w słyszalnej dla mnie odległości, podczas gdy nie śpię, co właściwie gwarantuje, że cię usłyszę. To, czy *zrozumiem* to, co mówisz, zależy od wielu rzeczy, ale nie wydaje się, żeby zależało od jakichkolwiek identyfikowalnych elementów wewnętrznej fenomenologii; żadne świadome przeżycie nie zagwarantuje tego, że cię zrozumiem bądź nie. Jeśli Sally wyobraża sobie swojego wujka Billa, w żaden sposób nie uniemożliwia jej to zrozumienia, że to wujek osoby mówiącej, nie jej wujek, zwolnił swojego prawnika; Sally *wie*, co osoba mówiąca miała na myśli; jedynie przypadkowo zajmuje się obrazem wujka Billa, wprowadzając niewielkie ryzyko zamieszania, gdyż zrozumienie osoby mówiącej w żadnym razie nie zależy od jej obrazów[18].

Zrozumienie nie może zatem zostać wyjaśnione poprzez powoływanie się na towarzyszącą mu fenomenologię, ale nie znaczy to, że jest ona w nim nieobecna. Szczególnie nie oznacza to, że model rozumienia, który ignoruje fenomenologię, odwoła się do zdroworozsądkowych intuicji na temat rozumienia. Głównym źródłem częstego sceptycyzmu w sprawie „maszynowego rozumienia" języka naturalnego jest z pewnością fakt, że takie systemy nigdy nie korzystają z czegoś w rodzaju „wizualnej" przestrzeni roboczej do rozbioru gramatycznego czy innej analizy informacji wejściowych. Gdyby tak było, poczucie, że rzeczywiście rozumieją to, co przetworzyły, byłoby o wiele większe (bez względu na to, czy nadal byłoby, jak sądzą niektórzy, iluzją). Obecnie, gdy komputer mówi „jasne" w odpowiedzi na informacje wejściowe, istnieje silna pokusa, aby zlekceważyć to stwierdzenie jako oczywiste oszustwo.

Ta pokusa jest istotnie interesująca. Na przykład trudno sobie wyobrazić, jak można zrozumieć pewne dowcipy bez uciekania się do obrazów umysłowych. Dwaj kumple siedzą w barze i piją; jeden odwraca się do drugiego i mówi: „Bud, myślę, że masz już dosyć – twoja twarz jest coraz bardziej rozmyta!". Czy nie stworzyliście jakiegoś obrazka czy przelotnego schematu, aby wyobrazić sobie błąd, który popełnił bohater dowcipu? To przeżycie jest dla nas przykładem tego, *czym jest zrozumienie czegoś*: napotykamy coś w pewien sposób kłopotliwego lub trudnego do rozszyfrowania, a przynajmniej na razie nieznanego – coś, co w taki czy inny sposób powoduje epistemiczne swędzenie, aż w końcu wykrzykniemy: *Aha! Zrozumiałem!* Pojawia się zrozumienie i dana sprawa zostaje przekształcona; staje się przydatna, jasna, kontrolowana przez ciebie. Przed momentem *t* na osi czasu była niezrozumiała; po momencie *t* została zrozumiana – jasno zaznaczona zmiana sytuacji, którą często można dokładnie określić w czasie, nawet jeśli jest to subiektywnie dostępna, introspekcyjnie odkryta przemiana. Jak

zobaczymy, jest błędem przyjmować ją jako model rozumienia, ale prawdą jest, że gdy początek rozumienia ma jakąkolwiek fenomenologię (gdy jesteśmy świadomi rozumienia czegoś), to jest to ta fenomenologia właśnie.

Coś w idei wyobrażeń umysłowych musi być prawdą, a jeśli nie należy pojmować jej w kategoriach „obrazów w głowie", będziemy musieli znaleźć coś lepszego. Obrazy umysłowe dotyczą wszystkich modalności, nie tylko wzroku. Wyobraź sobie kolędę *Cicha noc*, starając się jej nie śpiewać ani nie nucić. Czy mimo to „słyszysz" melodię uszami wyobraźni w konkretnej tonacji? Jeśli robisz tak jak ja, to słyszysz. Nie mam świetnego wyczucia muzycznego, więc nie potrafię powiedzieć „od wewnątrz", która to była tonacja, ale gdyby ktoś teraz zagrał na fortepianie *Cichą noc*, byłbym w stanie z dużą pewnością powiedzieć albo „Tak, to ta sama tonacja, w której słyszałem utwór", albo coś w rodzaju: „Nie, wyobrażałem ją sobie jakąś tercję wyżej"[19].

Nie tylko cicho mówimy sami do siebie, ale niekiedy robimy to w specyficznym „tonie". Czasem wydaje się też, że są tam słowa, ale nie słowa *słyszane*, a jeszcze kiedy indziej pojawiają się jedynie bardzo ulotne cienie czy ślady słów ucieleśniających nasze myśli. W czasach szczytu popularności psychologii introspekcyjnej burzliwie debatowano nad tym, czy istnieje coś takiego jak *całkowicie* „bezobrazowa" myśl. Teraz możemy tę kwestię pozostawić otwartą, nadmieniając jedynie, że wiele osób jest przeświadczonych, iż coś takiego istnieje, a inni z równą pewnością twierdzą, że nie. W następnym rozdziale opracujemy metodę rozwiązywania tego rodzaju konfliktów. Fenomenologia wyrazistych myśli nie ogranicza się do *mówienia* do siebie; możemy rysować dla siebie obrazy oczami wyobraźni, prowadzić dla siebie samochód ze skrzynią biegów, dotykać dla siebie jedwabiu lub smakować wyimaginowaną kanapkę z masłem orzechowym.

Bez względu na to, czy brytyjscy empiryści mieli rację, twierdząc, że wyobrażone (lub przypomniane) doznania są jedynie kopiami doznań oryginalnych „pochodzących z zewnątrz", mogą one dostarczyć nam przyjemności czy cierpienia tak samo jak „prawdziwe" wrażenia. Jak każdy marzyciel dobrze wie, fantazje erotyczne nie są być może całkowicie satysfakcjonującym substytutem rzeczywistej erotyki, ale są niewątpliwie czymś, za czym można by tęsknić, gdyby w jakiś sposób zostało odebrane. Dostarczają one nie tylko przyjemności; mogą wzbudzać prawdziwe doznania oraz inne, dobrze znane następstwa fizjologiczne. Możemy płakać, czytając smutną powieść, tak jak płakać może autor, gdy ją spisuje.

Wszyscy jesteśmy koneserami bólów i przyjemności wyobraźni, a wielu z nas uważa się za ekspertów w przygotowywaniu owych epizodów, które tak lubimy, ale nadal może nas zaskoczyć, jak niesłychana może to być umiejętność po poważnym wyszkoleniu. Uważam na przykład za niesamowite to, że podczas konkursów dla kompozytorów uczestnicy często nie dostarczają nagrań swoich utworów (czy występu na żywo); wysyłają zapis nutowy, a pewne siebie jury podejmuje ocenę *estetyczną* na podstawie odczytu nutowego oraz *słuchania muzyki w swoich głowach*. Jak dobre są najlepsze wyobraźnie muzyczne? Czy wytrenowany muzyk, sprawnie odczytując zapis, jest w stanie powiedzieć, jak zabrzmi nieskoordynowany dźwięk obojów i fletów bocznych na tle instrumentów strunowych? Istnieje mnóstwo anegdot na ten temat, lecz o ile mi wiadomo, jest to terytorium raczej niezbadane, czekające na mądre eksperymenty.

Wyobrażane doznania (jeśli tak możemy nazwać te fenomenologiczne detale) są podmiotem uznania i oceny estetycznej, ale dlaczego w takim razie rzeczywiste doznania liczą się o wiele bardziej? Dlaczego nie chcemy zadowolić się przypominanymi zachodami słońca czy oczekiwaniem na spaghetti z pesto? Większość przyjemności i bólu związanych z wydarzeniami w naszym życiu opiera się przecież na oczekiwaniu i przypominaniu. Czyste chwile

rzeczywistego doznania są tylko małą cząstką tego, co dla nas ważne. To, dlaczego – i w jaki sposób – coś jest dla nas ważne, będzie tematem kolejnych rozdziałów, ale fakt, że wyobrażone, przewidywane, przypominane wrażenia są inne od ulotnych doznań, może łatwo unaocznić nam kolejny niewielki eksperyment, co prowadzi nas do zagadnienia trzeciej części fenomu.

4. Afekt

Zamknij teraz oczy i wyobraź sobie, że ktoś bardzo mocno kopnął cię w lewą piszczel (jakieś trzydzieści centymetrów powyżej stopy), mając na sobie stalowo zakończony but. Wyobraź sobie przeszywający cię ból tak szczegółowo, jak potrafisz; wyobraź sobie, że łzy napływają ci do oczu, wyobraź sobie, że niemal mdlejesz, gdyż uderzenie bólu jest obrzydliwie silne i obezwładniające. Właśnie żywo to sobie wyobrażasz; czy czujesz jakiś ból? Czy możesz uczciwie poskarżyć mi się, że wypełnienie mojego polecenia doprowadziło cię do bólu? Okazuje się, że reakcje na to ćwiczenie są różne, ale nikt jeszcze nie oznajmił mi, że spowodowało ono prawdziwy ból. Dla niektórych jest ono niepokojące, dla innych to rozrywkowe ćwiczenie umysłu, które zdecydowanie nie jest tak nieprzyjemne, jak nawet najdelikatniejsze uszczypnięcie w ramię, godne miana „bólu".

Teraz wyobraź sobie, że taka sama scena kopnięcia w piszczel przyśniła ci się. Taki sen może być tak szokujący, że cię obudzi; może nawet okazać się, że przytulasz swoją goleń i pochlipujesz, a prawdziwe łzy pojawiają się w kącikach oczu. Nie byłoby jednak zaczerwienienia, śladu, siniaka, a po pełnym obudzeniu się i odzyskaniu orientacji od razu się przekonasz, że nie pozostał żaden ślad bólu twojej piszczeli – jeżeli w ogóle jakiś ból wystąpił. Czy bóle przyśnione to prawdziwe bóle, czy są tylko rodzajem wyobrażonego bólu? Czy czymś pomiędzy? A bóle wywołane podczas hipnozy?

Bóle przyśnione czy wywołane w hipnozie to stany umysłu, z których przynajmniej zdajemy sobie sprawę. Porównaj je za to ze stanami (umysłu?), które pojawiają się, gdy śpisz, kiedy przewracasz się na drugi bok i niechcący wykręcasz ramiona w dziwną pozycję, po czym, bez budzenia się, znów przewracasz się na bok, aby było ci wygodniej. Czy to są bóle? Na jawie stany wywołane takimi wygibasami byłyby bólami. Istnieją ludzie, na szczęście nieliczni, którzy urodzili się niewrażliwi na ból. Zanim zaczniesz im zazdrościć, musisz wiedzieć, że skoro nie poprawiają ułożenia ciała podczas snu (ale też na jawie!), szybko stają się kalekami, a ich stawy są zrujnowane przez bezustanne nadużycia, których nie tłumi żaden sygnał alarmowy. Takie osoby również oparzają się, tną i na różne inne sposoby skracają swoje nieszczęśliwe życie poprzez nieodpowiednie, opóźnione reagowanie (Cohen et al. 1955; Kirman et al. 1968).

Systemy alarmowe, czyli włókna bólu i powiązane z nimi obszary w mózgu, są bez wątpienia ewolucyjnym dobrodziejstwem, nawet jeśli oznaczają one włączanie się *niektórych* alarmów, z którymi niewiele możemy zrobić[20]. Ale dlaczego bóle muszą być tak *bolesne*? Dlaczego nie mógłby to na przykład być głośny dzwonek w uszach umysłu?

W takim razie jaką mamy korzyść, jeśli w ogóle, ze złości, strachu, nienawiści? (Zakładam, że ewolucyjnej korzyści z pożądania bronić nie trzeba). Jeszcze bardziej skomplikowany jest przypadek współczucia (*sympathy*). Angielskie słowo „sympathy" etymologicznie oznacza *współcierpienie*. Niemiecki odpowiednik to *Mitleid* (z bólem) oraz *Mitgefühl* (z czuciem). Pomyśl o *wibracji sympatycznej*, w której struna instrumentu muzycznego wydaje dźwięk, gdyż została poruszona przez wibracje innej struny, z którą jest związana tym, że dzielą częstotliwość rezonansową. Wyobraź sobie, że jesteś świadkiem głębokiego upokorzenia lub zawstydzenia twojego dziecka; nie możesz tego znieść: przepływają przez ciebie fale emocji zatapiające twoje myśli, wywracające twój spokój. Chcesz walczyć, płakać i coś uderzyć. Jest to

ekstremalny przykład współczucia. Dlaczego te zjawiska w nas zachodzą? I czym są?

To zainteresowanie adaptacyjnym znaczeniem (jeśli takowe w ogóle istnieje) rozmaitych stanów emocjonalnych będzie nam długo towarzyszyć w kolejnych rozdziałach. Teraz, podczas naszego spaceru, chciałbym zwrócić uwagę na niezaprzeczalną istotność odczuć wewnętrznych i emocji dla naszego przeświadczenia, że świadomość jest ważna. Weźmy na przykład *radość*. Wszystkie zwierzęta chcą *podtrzymać się przy życiu* – a przynajmniej bardzo się o to starają w większości warunków – ale zaledwie kilka gatunków zaskakuje nas umiejętnością *cieszenia się życiem* czy *bawienia się*. Przychodzą nam na myśl rozbrykane wydry ślizgające się na śniegu, bawiące się lwiątka, nasze koty i psy – ale nie pająki czy ryby. Konie, przynajmniej jako źrebięta, chyba cieszą się, że żyją, ale krowy czy owce zwykle są znudzone albo zobojętniałe. Czy nie zdarza się nam pomyśleć, że ptaki nie zasługują na latanie, gdyż nieliczne z nich, jeśli w ogóle jakieś, potrafią *doceniać* cudowność tej czynności? Radość nie jest pojęciem banalnym, ale o ile wiem, filozofowie nie poświęcili jej dotychczas wiele uwagi. Z pewnością nie zdobędziemy pełnego wyjaśnienia świadomości, dopóki nie uzasadnimy jej roli w naszym (i tylko naszym?) odczuwaniu radości. Jakie pytania należy zadać? Kolejny przykład pomoże nam zobaczyć, jakie wiążą się z tym trudności.

W Ameryce Południowej żyje gatunek ssaków naczelnych, który jest bardziej towarzyski niż inne i przejawia dziwne zachowanie. Członkowie tego gatunku często zbierają się w większych czy mniejszych grupach i podczas wspólnej paplaniny, w różnych okolicznościach, nakłaniają się do napadów mimowolnego konwulsyjnego oddychania, czegoś w rodzaju głośnego, niepohamowanego, wzajemnie umacnianego grupowego dyszenia, które czasem jest tak silne, że ich obezwładnia. Ataki nie wzbudzają jednak niechęci i wydaje się, że większość przedstawicieli tego gatunku dąży do nich, a niektórzy zdają się nawet od nich uzależnieni.

Może nas kusić myśl, że gdybyśmy tylko wiedzieli, jak to jest być nimi, tak od środka, zrozumielibyśmy to niespotykane zachowanie. Gdybyśmy mogli na nie spojrzeć „z ich perspektywy", wiedzielibyśmy, czemu służy. Jednak w tym przypadku możemy być raczej pewni, że taki dostęp i tak nie rozwiązałby zagadki. Dlatego, że ów dostęp już mamy; ten gatunek to *Homo sapiens* (który w rzeczy samej zamieszkuje między innymi Amerykę Południową), a owo zachowanie to śmiech[21].

Żadne inne zwierzę nie robi niczego podobnego. Biolog, który natknąłby się na tak wyjątkowe zjawisko, musiałby się przede wszystkim zastanowić, czemu ono służy, a nie znajdując żadnych przekonujących dowodów na płynące zeń bezpośrednie biologiczne korzyści, poczułby pokusę zinterpretowania tego nieproduktywnego zachowania jako ceny, którą gatunek musi płacić za inne dobrodziejstwo. Ale jakie? Cóż robimy lepiej dzięki mechanizmom, które, jako cenę wartą zapłacenia, niosą ze sobą podatność na nasze niemal uzależnienie od śmiechu? Czy śmiech w jakiś sposób pomaga „ulżyć stresowi" zbierającemu się w nas podczas skomplikowanych procesów myślowych związanych z naszym rozwiniętym życiem towarzyskim? Dlaczego jednak korzystam z *zabawnych* rzeczy, aby się odstresować? Dlaczego nie z *zielonych* rzeczy albo *zwykłych*, *płaskich* przedmiotów? I dlaczego właśnie to zachowanie jest produktem ubocznym odstresowywania się? Dlaczego nie chcemy stać w kółku i trząść się, bekać albo drapać sobie nawzajem plecy, nucić, wydmuchiwać nosy czy gorączkowo lizać swoje dłonie?

Zauważmy, że *widok od wewnątrz* jest dobrze znany i wcale nas nie dziwi. Śmiejemy się, *ponieważ jesteśmy rozbawieni*. Śmiejemy się, gdyż coś jest *śmieszne* – a śmiech jest *odpowiednią* reakcją na zabawne rzeczy, w przeciwieństwie do lizania dłoni. Jest oczywiste (a wręcz *zbyt* oczywiste), dlaczego się śmiejemy. Śmiejemy się ze szczęścia i zachwytu, z radości oraz dlatego, że coś jest komiczne. Jeśli istnieje jakaś *virtus dormitiva* w wyjaśnieniu, oto ona: śmiejemy się

z powodu komiczności bodźca[22]. Jest to oczywiście prawda; tylko dlatego się śmiejemy, gdy śmiejemy się szczerze. Wesołość jest podstawowym elementem prawdziwego śmiechu. Tak jak ból jest podstawowym elementem nieudawanej reakcji na ból. Nie możemy temu zaprzeczać, gdyż jest to oczywista prawda.

Potrzebujemy jednak wyjaśnienia śmiechu, które wychodzi ponad tę oczywistą prawdę, tak jak standardowe objaśnienia bólu wychodzą ponad oczywistość. Możemy podać zupełnie rozsądne biologiczne wyjaśnienie istnienia bólu (w rzeczy samej właśnie je zarysowaliśmy); chcemy podobnego, rozsądnego wytłumaczenia tego, dlaczego istnieje wesołość i śmiech.

Wiemy natomiast z góry, że nawet jeśli uda nam się takie wytłumaczenie podać, to nie usatysfakcjonuje ono wszystkich! Ludzie, którzy uważają się za *antyredukcjonistów*, narzekają, że biologiczne wyjaśnienie bólu i zachowań bólowych *pomija bolesność*, pomija „wewnętrzną własność bolesności", która sprawia, iż ból jest tym, czym jest, i te osoby prawdopodobnie zarzucą nam to samo przy jakiejkolwiek próbie objaśnienia śmiechu: pomija ona wewnętrzną własność zabawności. Jest to standardowy zarzut wobec tego rodzaju wyjaśnień: „Jedyne, co wyjaśniłeś, to związane z tym zjawiskiem *zachowania* i *mechanizmy*, ale pominąłeś rzecz samą w sobie, czyli ból i jego okropność". Pojawiają się tu skomplikowane pytania, które zostaną szczegółowo zanalizowane w rozdziale 12, teraz jednak możemy zwrócić uwagę na fakt, że każde wyjaśnienie bólu, które *uwzględni* okropność, będzie koliste – będzie miało *virtus dormitiva* nie do pominięcia. Analogicznie, każde poprawne ujęcie śmiechu *musi* pominąć domniemanie wpisanej weń zabawności, entuzjazmu czy śmieszności, gdyż ich obecność jedynie odwiodłaby nas od zamiaru odpowiedzi na to pytanie.

Fenomenologia śmiechu jest hermetycznie zamknięta; my *po prostu widzimy* bezpośrednio, naturalnie, bez konieczności wyciągania wniosków, z oczywistością przekraczającą „intuicję", że śmiech to coś, co wiąże się z zabawnością – jest to „odpowiednia" reakcja na śmieszność. Niewykluczone, że trochę udało nam się to zanalizować: odpowiednia reakcja na coś śmiesznego to rozbawienie (wewnętrzny stan umysłu); naturalnym wyrażeniem rozbawienia (gdy nie musimy go powstrzymywać czy ukrywać, jak to czasem bywa) jest śmiech. Wydaje się, że mamy tu coś, co naukowiec nazwałby zmienną pośredniczącą – rozbawienie – pomiędzy bodźcem a reakcją. Zmienna zdaje się w istotny sposób powiązana z obiema stronami. Rozbawienie jest z definicji tym, co powoduje szczery śmiech, i również z definicji jest czymś, co jest spowodowane przez coś śmiesznego. Wszystko to jest oczywiste i wydawałoby się, że nie ma potrzeby dalszego wyjaśniania. Jak powiedział Wittgenstein, wyjaśnienia muszą się gdzieś skończyć. Jednak to, co naprawdę mamy przed sobą, to surowy – ale z pewnością możliwy do wyjaśnienia – fakt dotyczący ludzkiej psychologii. Musimy wykroczyć poza czystą fenomenologię, jeśli chcemy wyjaśnić jakąkolwiek kwestię z ogrodu fenomenologicznego.

Podane przykłady fenomenologiczne w swojej różnorodności mają bodaj jedną cechę wspólną. Z jednej strony są to nasi najbliżsi znajomi; nie istnieje nic, co moglibyśmy znać lepiej niż elementy naszej osobistej fenomenologii – a przynajmniej tak mogłoby się wydawać. Z drugiej strony są one niedostępne dla nauki materialistycznej; nic nie mogłoby być mniej podobne do elektronu, cząsteczki czy neuronu niż *to, jak wygląda dla mnie ten zachód słońca* – a przynajmniej tak mogłoby się wydawać. Filozofowie pozostają rzeczywiście pod wrażeniem obu tych cech i wielokrotnie podkreślali ich problematyczne aspekty. Dla niektórych wielką zagadką jest owa bliskość: Jak możemy być tak *niekorygowalni* lub mieć *uprzywilejowany dostęp* czy też *bezpośrednio pojmować* te kwestie? Jaka jest różnica pomiędzy naszymi epistemicznymi relacjami z naszą fenomenologią a epistemicznymi relacjami z obiektami ze świata zewnętrznego? Dla innych wielką zagadką są niezwykłe „właściwości wewnętrzne" – lub po łacinie *qualia* – naszej fenomenologii: W jaki sposób coś stworzone z cząstek materialnych może

być radością, której doświadczam, lub cechować się „ostateczną jednorodnością” (Sellars 1963/1991) różowej kostki lodu, którą właśnie sobie wyobrażam, czy też znaczyć dla mnie tyle, co ból, który odczuwam?

Odnalezienie materialistycznego wyjaśnienia, które oddaje sprawiedliwość wszystkim tym zjawiskom, nie będzie łatwym zadaniem. Poczyniliśmy jednak pewien postęp. Nasz krótki spis zawiera pewne przykłady, w których odrobina wiedzy o rządzących nimi mechanizmach podważa – a może nawet uzurpuje – autorytet, który zwykle udzielamy temu, co jest oczywiste przy analizie stanów psychicznych. Poprzez przysuwanie się bliżej niż zwykle do eksponatów oraz poprzez spoglądanie na nie pod różnym kątem zaczynamy niweczyć zły urok, uwalniać się od „magii” w ogrodzie fenomenologicznym.

Rozdział 4

Metoda dla fenomenologii

1. Pierwsza osoba liczby mnogiej

Poważnej zoologii nie uprawia się, jedynie spacerując po zoo, dostrzegając to i owo czy zachwycając się ciekawostkami. Poważna zoologia wymaga precyzji, która opiera się na wspólnie ustalonych metodach opisu i analizy, aby inni zoologowie mieli pewność, co masz na myśli. Poważna fenomenologia potrzebuje jeszcze jaśniejszych, neutralnych metod opisu, gdyż wydaje się, że nawet dwóch ludzi nie używa jednego słowa w ten sam sposób, a przecież każdy z nich jest znawcą języka. Niewiarygodne jest to, jak często „akademickie" dyskusje dotyczące kontrowersyjnych kwestii fenomenologicznych kończą się uderzaniem pięściami w biurko i kakofonią przekrzykujących się głosów. Jest to tym bardziej zaskakujące, że według wielowiekowej filozoficznej tradycji *wszyscy zgadzamy się* co do tego, co znajdujemy, gdy „spoglądamy do wewnątrz" w naszą własną fenomenologię.

Uprawianie fenomenologii zwykle wydaje się rzetelną praktyką społeczną, kwestią tworzenia wspólnych obserwacji. Gdy Kartezjusz napisał *Medytacje* w postaci monologu z samym sobą, oczekiwał od swoich czytelników, że przystaną na jego obserwacje, ale przemierzając w swoich umysłach dokładnie opisaną przez niego drogę i otrzymując te same rezultaty. Identycznie postępowali brytyjscy empiryści, Locke, Berkeley i Hume, zakładając, że dokonują *introspekcji*, która może być łatwo powtórzona przez ich czytelników. Locke przyjął to założenie w *Rozważaniach dotyczących rozumu ludzkiego* (1690), nazywając swoją metodę „prostą metodą historyczną" – żadnych zawiłych dedukcji czy teoretyzowania *a priori*, jedynie ustalanie zaobserwowanych faktów oraz przypominanie czytelnikom, co jest oczywiste dla każdego obserwatora. W zasadzie niemal każdy autor piszący o świadomości wprowadzał coś, co moglibyśmy nazwać *domniemaniem pierwszej osoby liczby mnogiej*: pomimo tajemnic skrywających się w świadomości *my* (ty, łagodny czytelniku, i ja) możemy spokojnie rozmawiać o naszych wspólnych znajomych, czyli o tym, co wszyscy znajdujemy w swoich strumieniach świadomości. Okazuje się, że poza paroma krzykliwymi wyjątkami czytelnicy zawsze uczestniczyli w tym spisku.

Wszystko byłoby w porządku, gdyby nie wstydliwy fakt, że spory i sprzeczności niweczą ustalenia zawarte w warunkach uprzejmej, obopólnej zgody. Sami siebie oszukujemy w pewnej kwestii. Być może chodzi o założenie, że w dużym stopniu jesteśmy do siebie, ogólnie rzecz biorąc, podobni. Być może, gdy poznajemy różne szkoły fenomenologiczne, przyłączamy się do tej, która wydaje się nam prawdziwa, natomiast żadna szkoła opisu fenomenologicznego na ogół nie myli się co do życia wewnętrznego swoich członków, czyni zaś niewinne, choć zbyt szerokie uogólnienia, stawiając niepoparte niczym tezy o tym, co się dzieje w umyśle każdego człowieka.

Niewykluczone też, że oszukujemy samych siebie co do niezawodności introspekcji, czyli naszych własnych zdolności samoobserwacji w świadomym umyśle. Od czasów Kartezjusza i *cogito ergo sum* owa umiejętność była postrzegana jako coś, co nie jest podatne na błąd; mamy uprzywilejowany dostęp do naszych własnych myśli i uczuć, dostęp, który gwarantuje nam lepszy wgląd niż komukolwiek z zewnątrz. („Wyobraź sobie, że ktoś próbuje ci wmówić, że mylisz się co do swoich myśli i odczuć!") Jesteśmy albo „nieomylni" – czyli zawsze mamy rację

– albo przynajmniej „niekorygowalni” – czy się mylimy, czy nie, nikt inny nie może nas poprawić (Rorty 1970).

Niewykluczone jednak, że ta doktryna nieomylności, pomimo swojego silnego ugruntowania, jest błędna. Nawet jeśli *jesteśmy*, ogólnie rzecz biorąc, tacy sami, być może niektórzy obserwatorzy mylą się w opisach, a jednocześnie są tak pewni swych racji, że trudno ich przekonać do jakichkolwiek poprawek. (Są niekorygowalni w sensie pejoratywnym). Wówczas pojawia się niezgoda. Poza tym istnieje jeszcze jedna możliwość i myślę, że jest ona o wiele bliższa prawdy: oszukujemy samych siebie, gdyż uważamy, że „introspekcja” jest *jedynie* kwestią „spoglądania i dostrzegania”. Podejrzewam, że gdy uważamy, iż korzystamy jedynie z mocy wewnętrznej *obserwacji*, zawsze bierzemy się za pewnego rodzaju improwizowane *teoretyzowanie* – a jesteśmy nadzwyczaj naiwnymi teoretykami, właśnie dlatego, że tak *mało* mamy do „obserwacji”, a tak dużo do oznajmienia, bez strachu przed popadnięciem w sprzeczność. Gdy wspólnie zaglądamy do wnętrza, jesteśmy niczym legendarni ślepcy, którzy badają odmienne części ciała słonia. Z początku pomysł ten wydaje się absurdalny, ale spójrzmy, co możemy ustalić w jego obronie.

Czy zaskoczyło cię coś, o czym mówiliśmy podczas naszej wycieczki po fenomie w poprzednim rozdziale? Czy na przykład zdziwiło cię, że nie potrafisz zidentyfikować karty, zanim pojawiła się niemal dokładnie przed tobą? Zauważyłem, że większość ludzi jest zaskoczona – nawet ci, którzy wiedzą o ograniczonej ostrości widzenia peryferyjnego. Jeśli cię to zdziwiło, musi to oznaczać, że przed tą zaskakującą demonstracją najprawdopodobniej twoje wypowiedzi na ten temat byłyby błędne. Ludzie często twierdzą, że mają bezpośrednią znajomość większej zawartości pola widzenia peryferyjnego niż w rzeczywistości. Skąd te twierdzenia? Nie biorą się stąd, że osoby te bezpośrednio i niepoprawnie zaobserwowały taką możliwość, ale dlatego, że wydaje się to *racjonalne*. W normalnych warunkach nie dostrzegasz przecież żadnych wielkich luk w swoim polu widzenia, a z pewnością, gdyby był tam obszar bez koloru, widać byłoby różnicę, a poza tym, gdziekolwiek nie spojrzysz, wszystko ma kolory i szczegóły. Jeśli myślisz, że twoje subiektywne pole widzenia jest wewnętrznym obrazem skomponowanym z kolorów i kształtów, to przecież jest racjonalne, że każdy fragment płótna musi być pomalowany na *jakiś* kolor – nawet puste płótno ma *jakiś* kolor! Jest to jednak wniosek wyciągnięty z wątpliwego modelu twojego subiektywnego pola widzenia, a nie z czegoś zaobserwowanego bezpośrednio.

Czy twierdzę, że nie mamy w ogóle żadnego uprzywilejowanego dostępu do przeżyć świadomych? Nie, ale jestem zdania, że zwykle wydaje się nam, iż jesteśmy o wiele bardziej odporni na błąd niż w rzeczywistości. Gdy w ten sposób kwestionuje się uprzywilejowany dostęp, ludzie zwykle przyznają, że nie mają szczególnego wglądu w *przyczyny i skutki* swoich świadomych przeżyć. Mogą być na przykład zaskoczeni, że smakują nosem i słyszą nuty basowe swoimi stopami, ale nigdy nie twierdzili, że są autorytetem w kwestii przyczyn i skutków swoich przeżyć. Są jedynie autorytetami, jak mówią, w kwestii przeżyć samych w sobie, w oderwaniu od przyczyn i skutków. Nawet jeśli ludzie *twierdzą*, że są autorytetami tylko w kwestii treści przeżyć, a nie ich przyczyn i skutków, często przekraczają granice, które sami sobie wyznaczyli. Czy na przykład zgodzisz się z następującymi sądami? (Zmyśliłem przynajmniej jeden z nich).

(1) Możesz mieć przeżycie plamy, która jest czerwona i zielona na całej powierzchni w tym samym momencie – plamy, która jest w *obu* kolorach (niezmieszanych) w tym samym czasie.

(2) Gdy spojrzysz na żółte koło na niebieskim tle (przy dobrym oświetleniu), a jasność oraz nasycenie żółtego i niebieskiego zostaną ustawione tak, aby były sobie równe, granica

między kolorami zniknie.

(3) Istnieje dźwięk zwany „dźwiękiem Sheparda", który pozornie wciąż wzrasta, ale nigdy tak naprawdę nie staje się wyższy.

(4) Istnieje zioło, którego przedawkowanie powoduje, że nie można zrozumieć własnego mówionego języka rodzimego. Dopóki zioło nie przestanie działać, słuch jest nienaruszony, a słyszane dźwięki nie są rozmyte, nie słyszy się też żadnych dźwięków dodatkowych, ale słowa, które do ciebie docierają, brzmią jak zupełnie obcy język, chociaż w pewnym sensie wiesz, że obce nie są.

(5) Gdy masz zasłonięte oczy i do twojego ramienia zostanie przyłożony wibrator, a jednocześnie dotykasz nosa, poczujesz, że nos rośnie ci jak u Pinokia; jeśli wibrator zostanie przesunięty w inne miejsce, będziesz miał dziwne uczucie wypychania nosa na drugą stronę i poczujesz, że twój palec zatrzymuje się gdzieś wewnątrz czaszki.

Wymyśliłem sąd numer 4, ale teoretycznie mógłby być prawdziwy. W szeroko badanym przypadku neuropatologicznym, zwanym prozopagnozją, wzrok pozostaje nienaruszony i bez problemu rozpoznaje się większość rzeczy, ale twarze najbliższych przyjaciół i znajomych są nierozpoznawane[23]. Chcę w ten sposób znów podkreślić nie tyle brak dostępu do natury czy treści twoich świadomych przeżyć, ile to, że po prostu powinniśmy uważać na *bardzo* kuszącą zbytnią pewność siebie w tym zakresie.

Podczas wycieczki z przewodnikiem po fenomie zaproponowałem wiele prostych eksperymentów. Nie było to w duchu „czystej" fenomenologii. Fenomenolodzy często twierdzą, że ponieważ nie jesteśmy autorytetami w kwestii fizjologicznych przyczyn i skutków naszej fenomenologii, powinniśmy je zignorować, gdy próbujemy stworzyć czysty, neutralny, preteoretyczny opis tego, co uważamy za „dane" nam w codziennym doświadczeniu. Być może, ale pomyśl, ilu ciekawych mieszkańców fenomu nigdy byśmy wówczas nie poznali! Zoolog, który próbowałby wyciągać wnioski dla całej nauki jedynie z obserwacji psa, kota, konia, rudzika i złotej rybki, prawdopodobnie paru rzeczy by nie uwzględnił.

2. Perspektywa trzecioosobowa

Jako że mamy uprawiać fenomenologię *nieczystą*, w kwestii metodologii musimy być ostrożniejsi. Fenomenologowie zwykle przyjmują *perspektywę pierwszoosobową* Kartezjusza, w której *ja* opisuję w monologu (który pozwalam *tobie* usłyszeć), co *ja* odnajduję w *moim* świadomym przeżyciu, zakładając, że *my* się ze sobą zgodzimy. Próbuję jednak pokazać, że owo wygodne współdziałanie wynikające z perspektywy *pierwszoosobowej liczby mnogiej* to zdradliwy inkubator błędów. W historii psychologii rozpoznanie tego metodologicznego problemu doprowadziło do upadku introspekcjonizmu i pojawienia się behawioryzmu. Behawioryści bardzo starali się unikać spekulacji dotyczących tego, co się dzieje w *moim* umyśle, w *twoim* umyśle, w *jego* czy *jej* umyśle. W efekcie wytworzyli *perspektywę trzecioosobową*, gdzie jedynie fakty zebrane „z zewnątrz" są uznawane za dane. Możesz sfilmować ludzi w akcji, a następnie zmierzyć współczynnik błędu w zadaniach wymagających ruchu, czasy reakcji podczas naciskania przycisków czy poruszania dźwigni, fale mózgowe, ruchy gałek ocznych, rumienienie się (jeżeli masz maszynę, która obiektywnie to mierzy) oraz reakcję skórno-galwaniczną (przewodnictwo elektryczne wykrywane przez „wykrywacze

kłamstw"). Możesz otworzyć czaszki badanych (chirurgicznie bądź przez urządzenia obrazujące mózg), aby zobaczyć, co się dzieje w ich *mózgach*, ale nie możesz *zakładać*, że coś dzieje się w ich *umysłach*, gdyż z umysłów nie możemy otrzymać żadnych danych, jeśli korzystamy z intersubiektywnie weryfikowalnych metod nauk fizycznych.

Najprościej to ujmując, behawioryści uważali, że skoro nigdy nie można „bezpośrednio spojrzeć" w umysł człowieka, a jedynie trzeba brać za dobrą monetę, co człowiek mówi, to jakiekolwiek fakty dotyczące zdarzeń umysłowych nie zaliczają się do danych naukowych, gdyż nigdy nie będą mogły być odpowiednio zweryfikowane metodami obiektywnymi. Owa *metodologiczna* wątpliwość, która jest zasadą rządzącą dziś *wszystkimi* eksperymentami psychologicznymi i neuronaukowymi (nie tylko w badaniach „behawiorystycznych"), jest zbyt często podnoszona do rangi takiej czy innej *ideologicznej* zasady, na przykład:

Zdarzenia umysłowe nie istnieją. (I kropka! – to tak zwany „bosy behawioryzm").

Zdarzenia umysłowe istnieją, ale nie mają na nic wpływu, więc nauka nie może ich badać (epifenomenalizm – zobacz rozdział 12, sekcja 5).

Zdarzenia umysłowe istnieją i mają na nas wpływ, ale owe wpływy nie mogą być badane przez naukę, która musi zadowolić się teoriami „peryferyjnych" czy też „niższych" wpływów i procesów w mózgu. (Ten pogląd jest dość popularny wśród neuronaukowców, zwłaszcza tych, którzy mają wątpliwości co do „teoretyzowania". W rzeczywistości jest to dualizm; ci badacze najwyraźniej zgadzają się z Kartezjuszem, że umysł to nie mózg, i są gotowi zadowolić się jedynie teorią mózgu).

Te poglądy prowadzą od jednego bezzasadnego wniosku do drugiego. Nawet jeśli zdarzenia umysłowe nie są zaliczane do *danych* naukowych, nie oznacza to, że nie możemy naukowo ich badać. Czarne dziury i geny nie są danymi naukowymi, ale rozwinęliśmy dobre naukowe teorie, aby je badać. Wyzwaniem jest stworzenie teorii zdarzeń umysłowych opartej na danych uznanych przez metodę naukową.

Taka teoria będzie musiała być skonstruowana z trzecioosobowego punktu widzenia, gdyż *cała* nauka opiera się na tej perspektywie. Niektórzy powiedzą ci, że taka teoria świadomego umysłu jest niemożliwa. Szczególnie filozof Thomas Nagel twierdzi, że:

[I]stnieją jednak pewne aspekty świata, życia i nas samych, których nie można adekwatnie zrozumieć z najobiektywniejszego punktu widzenia, i to bez względu na to, jak bardzo taka perspektywa wzbogaciła nasze poznanie. Wiele spraw łączy się w istotny sposób z określonym punktem widzenia albo rodzajem punktu widzenia. Próba całkowitego wyjaśnienia świata w obiektywnych terminach, w oderwaniu od tych perspektyw, prowadzi w nieunikniony sposób do fałszywych redukcji albo do oczywistego odrzucania jakichś zjawisk, których istnienie nie budzi wątpliwości. [Nagel 1986/1997, s. 11–12]

Zobaczymy. Przedwczesna jest dyskusja o tym, co może lub czego nie może objaśnić teoria, jeśli nie wiemy, o czym tak naprawdę mówi. Jeśli jednak chcemy uczciwie jej wysłuchać, w świetle takiego sceptycyzmu będziemy musieli znaleźć neutralny sposób *opisu danych* – sposób, który nie przesądza tej kwestii. Wbrew pozorom istnieje taka neutralna metoda, którą najpierw opiszę, a następnie przyjmę.

3. Heterofenomenologia[24]

Termin ten brzmi złowieszczo; nie tylko fenomenologia, ale *hetero*fenomenologia. Cóż może on oznaczać? Jest to w rzeczywistości coś znanego nam wszystkim, laikom i naukowcom, coś, co musimy jednak przedstawić z fanatyczną ostrożnością, wyraźnie zaznaczając, co zakłada i co z niego wynika, gdyż wiąże się to z ważnym krokiem teoretycznym. Omijając wszelkie

kuszące skróty, przedstawiam *neutralną* ścieżkę prowadzącą od obiektywnej nauki fizycznej i trzecioosobowego punktu widzenia do metody opisu fenomenologicznego, który może (w zasadzie) oddać sprawiedliwość najbardziej intymnym i niewyrażalnym subiektywnym przeżyciom – ani przez moment nie porzucając metodologicznych wymagań nauki.

Chcemy mieć teorię świadomości, ale można się spierać, które istoty są świadome. Czy świadome są noworodki? A żaby? A co z ostrygami, mrówkami, roślinami, robotami, zombi...? Na razie powinniśmy pozostać w tej kwestii neutralni, natomiast istnieje klasa istot, którą powszechnie uznaje się za obdarzoną świadomością, a są to dorosłe istoty ludzkie.

Niektóre z tych dorosłych istot ludzkich *mogą* być zombi – w „teoretycznym" sensie filozoficznym. Pojęcie *zombi* pochodzi z tradycji wudu z Haiti i odnosi się, w jej kontekście, do „żywego trupa", osoby ukaranej za jakiś zły czyn i skazanej na szuranie nogami, mamrotanie i wpatrywanie się przed siebie martwymi oczami, bezmyślnie wypełniającej polecenia jakiegoś kapłana czy szamana wudu. Wszyscy widzieliśmy zombi w horrorach i wiemy, że można je bez trudu odróżnić od normalnych ludzi. (Ogólnie rzecz biorąc, haitańskie zombi nie potrafią tańczyć, opowiadać dowcipów, żywo dyskutować o filozofii, bronić swojego zdania w dowcipnej konwersacji – i po prostu wyglądają obrzydliwie[25]). Filozofowie używają jednak pojęcia „zombi" na oznaczenie innej kategorii wyimaginowanych istot ludzkich. Według terminologicznej konwencji filozofów zombi jest lub byłby człowiekiem, który wykazuje się zupełnie naturalnym zachowaniem, jest bystry, rozmowny i pełen życia, ale w rzeczywistości nie jest wcale świadomy, tylko jest czymś w rodzaju automatu. Cały sens tego filozoficznego pojęcia jest taki, że nie potrafimy odróżnić zombi od normalnego człowieka, badając jedynie ich zewnętrzne zachowanie. A jedynie takie zachowanie możemy obserwować u przyjaciół i sąsiadów, tak więc *niektórzy z twoich najlepszych przyjaciół mogą być zombi*. Jest to jednak tradycja, wobec której muszę początkowo pozostać neutralny. Opisywana metoda nie zakłada nic na temat *rzeczywistej świadomości* z pozoru normalnych dorosłych ludzi, ale to na nich się koncentruje, ponieważ jeśli świadomość gdziekolwiek występuje, to właśnie w tych osobnikach. W momencie gdy ustalimy, jak mógłby wyglądać zarys teorii świadomości ludzkiej, będziemy mogli skupić się na świadomości (lub jej braku) u innych gatunków, takich jak szympansy, delfiny, rośliny, zombi, Marsjanie oraz tostery (eksperymenty myślowe często pobudzają filozofów do fantazjowania).

Dorosłe istoty ludzkie (odtąd będę pisać po prostu o ludziach) są badane w wielu dziedzinach nauki. Ich ciała są analizowane przez biologów, lekarzy, dietetyków czy inżynierów (którzy zadają pytania typu: Jak szybko człowiek może pisać na klawiaturze? Jak wytrzymały na rozciąganie jest ludzki włos?). Są również badani przez psychologów i neuronaukowców, którzy umieszczają pojedyncze osoby, zwane *badanymi*, w różnych eksperymentalnych sytuacjach. W większości eksperymentów osoby badane muszą zostać najpierw skategoryzowane i przygotowane. Trzeba ustalić nie tylko, ile mają lat, jakiej są płci, czy są lewo-, czy praworęczne, jak są wykształcone i tak dalej, ale trzeba im także powiedzieć, *co mają robić*. Jest to najbardziej uderzająca różnica pomiędzy osobami badanymi i na przykład kulturami bakterii biologa, próbkami egzotycznych minerałów inżyniera, roztworami chemika czy szczurami, kotami i gołębiami psychologa zwierząt.

Ludzie są jedynymi przedmiotami badań naukowych, których przygotowanie zwykle (choć nie zawsze) zakłada komunikację werbalną. Jest to w pewnym sensie kwestia etyki w nauce: ludzie nie mogą zostać wykorzystani w eksperymentach bez własnej zgody, a otrzymanie takiej zgody bez interakcji werbalnej nie jest możliwe. Z naszego punktu widzenia ważniejszy jest jednak fakt, iż komunikację werbalną stosuje się do przygotowywania i przeprowadzania eksperymentów. Osoby badane są proszone o wykonanie różnych czynności

intelektualnych, o rozwiązanie problemu, o odszukanie czegoś na ekranie, o naciskanie przycisków, ocenianie i tak dalej. Wiarygodność większości eksperymentów zależy od tego, czy owo przygotowanie jest jednolite i skuteczne. Gdy na przykład okaże się, że polecenia były wydawane w języku tureckim, a jedynym językiem osób badanych był angielski, niepowodzenie eksperymentu jest właściwie gwarantowane. Oznaka nawet najmniejszego niezrozumienia może narazić cały eksperyment na niepowodzenie, jest więc kwestią istotną, aby ta praktyka przygotowywania badanych przez komunikację werbalną była zatwierdzona.

Cóż wiąże się z ową praktyką mówienia do osób badanych? Jest to element nie do pominięcia w eksperymentach psychologicznych, ale czy zakłada u badanych świadomość? Czy eksperymentatorzy nie cofają się na pozycje introspekcjonistów, musząc wierzyć niesprawdzonym słowom badanych, że rozumieją instrukcje? Czy nie ryzykujemy oszukania przez zombi, roboty czy innych mistyfikatorów?

Musimy bliżej przyjrzeć się szczegółom typowego eksperymentu z osobą badaną. Załóżmy, że, jak to się zwykle dzieje, powstaje wiele zapisów eksperymentu: taśma wideo, taśma dźwiękowa, encefalograf itp. Nic, co nie jest nagrane, nie zostanie uznane za dane. Spójrzmy na nagrania dźwiękowe – głównie mowy – badanych oraz eksperymentatorów. Dźwięki wydawane przez osoby badane są wytworzone środkami fizycznymi, więc w zasadzie można je wyjaśnić i przewidzieć na gruncie fizyki, z użyciem tych samych zasad, praw i modeli, z których korzystamy, aby wyjaśnić i przewidzieć hałasy czy uderzenia dobywające się z silnika samochodowego. Innymi słowy, skoro dźwięki powstają fizycznie, moglibyśmy dodać zasady fizjologii i wyjaśniać dźwięki, korzystając z zasobów tej nauki, tak samo jak wyjaśniamy beknięcia, chrapanie, burczenie w brzuchu i strzykanie w kościach. Interesują nas jednak głównie dźwięki głosu, a dokładniej pewien ich podzbiór (ignorujemy okazjonalne czknięcia, kichnięcia czy ziewnięcia), który *wydaje się* podatny na analizę składniową i semantyczną. Nie zawsze jest jasne, które dźwięki należy umieścić w tym podzbiorze, ale istnieje bezpieczny sposób: dajemy kopie nagrań trzem stenografom i prosimy, aby przygotowali *transkrypcje* surowych danych.

Ten prosty krok ma wiele następstw; razem z nim przenosimy się z jednego świata – świata czystych fizycznych dźwięków – do innego: świata słów i znaczeń, składni i semantyki. Krok ten niesie ze sobą radykalną rekonstrukcję danych, abstrakcję daleką od akustycznych i innych fizycznych cech, tworzącą ciąg słów (choć wciąż uzupełnionych precyzyjnym określeniem czasu – zobacz na przykład Ericsson i Simon 1984). Co rządzi tą rekonstrukcją? Mimo że przypuszczalnie istnieją regularne i możliwe do odkrycia związki między właściwościami fizycznymi fali akustycznej zapisanej na taśmie oraz *fonemami*, które słyszy stenograf, nie wiemy jeszcze o nich na tyle dużo, aby opisać je szczegółowo. (Gdybyśmy wiedzieli, problem skonstruowania urządzenia, które spisywałoby podyktowany tekst, byłby rozwiązany. W tej kwestii postęp był całkiem duży, ale wiele pozostaje jeszcze do dopracowania). Czekając na wyniki owych badań z zakresu akustyki i fonologii, możemy nadal ufać naszym transkrypcjom jako obiektywnym przedstawieniom danych, jeśli tylko będziemy pamiętać o kilku podstawowych środkach ostrożności. Po pierwsze, proszenie stenografów o sporządzenie transkrypcji (zamiast powierzania tego zadania na przykład eksperymentatorowi) zapewnia nam ochronę zarówno przed zamierzoną, jak i bezwiedną stronniczością czy nadinterpretacją. (Stenografowie sądowi spełniają tę samą, neutralną rolę). Przygotowanie trzech niezależnych transkrypcji pokazuje nam, w jakim stopniu ten proces jest obiektywny. Jeśli nagranie jest dobrej jakości, transkrypcje prawdopodobnie będą się ze sobą zgadzały co do słowa, z wyjątkiem maleńkiej części, jednego procenta słów. Tam, gdzie nie będą się one ze sobą zgadzały, jeśli chcemy, możemy po prostu pozbyć się danych lub przystać na zgadzające się ze sobą dwie wersje zapisu.

Taka transkrypcja nie tworzy, ściśle rzecz biorąc, oczywistych danych, gdyż, jak widzieliśmy, powstaje przez poddawanie surowych danych procesowi interpretacji. Proces ten zależy od założenia dotyczącego języka, jakim posługuje się osoba badana, oraz od pewnych jej intencji. Aby dostrzec tę różnicę, porównaj zadanie dane stenografom z zadaniem stworzenia transkrypcji nagrań śpiewu ptaków czy chrumkania świń. Gdy człowiek mówi: „Moe nacisnoć pszycis lewom renkom?", wszyscy stenografowie zgodzą się, że zapytał: „Mogę nacisnąć przycisk lewą ręką?" – a to dlatego, że znają język polski i właśnie to zdanie ma sens w tym kontekście. Jeśli natomiast badany powie: „Teraz kropka porusza się z dóry na gół", pozwolimy stenografom wprowadzić poprawkę i napisać: „Teraz kropka porusza się z góry na dół". Nie mamy dostępnej tego typu strategii oczyszczającej dla śpiewu ptaków czy chrumkania świń – przynajmniej dopóki jakiś badacz nie odkryje, że istnieją pewne normy w tych dźwiękach, i nie stworzy systemu ich opisu.

Bez żadnego wysiłku – a właściwie mimowolnie – „rozumiemy" potok dźwięków w procesie zamieniania go na słowa. (Lepiej pozwólmy stenografom zmienić „z dóry na gół" na „z góry na dół", gdyż prawdopodobnie i tak zrobią to nieświadomie). Proces ten jest jednocześnie wysoce wiarygodny, jak i zupełnie niezauważalny, co nie powinno jednak przysłaniać nam faktu, że jest on skomplikowany nawet wówczas, gdy nie docieramy do momentu samego zrozumienia, lecz zatrzymujmy się na etapie rozpoznawania słów. Kiedy stenograf notuje: „Dla mnie w tym złym przeczuciu był uderzający rodzaj oddalenia, kusząca nuta przedsmaku i zniewagi, rozmaite przewidujące potwierdzenia, które ujawniały powierzchnię pod kolejną powierzchnią", może nie mieć najmniejszego pojęcia, co to znaczy, a jednak będzie pewien, że te słowa są słowami, które mówiący rzeczywiście chciał wypowiedzieć, i że udało mu się to, cokolwiek miał na myśli.

Ale zawsze istnieje możliwość, że mówiący również nie wiedział, co jego słowa miałyby znaczyć. W końcu osoba badana *mogłaby* być zombi czy papugą w ludzkim stroju, czy też komputerem wyposażonym w syntezator mowy. Bądź też, bardziej przyziemnie, osoba badana mogła być zdezorientowana, pochwycić jakąś błędnie rozumianą teorię, a wręcz żartować sobie z eksperymentatora, wyrzucając z siebie potok bezsensownych słów. W tym momencie proces tworzenia transkrypcji na podstawie danych jest neutralny w stosunku do wszystkich tych osobliwych możliwości, nawet jeśli przebiega z metodologicznym założeniem, że z nagrania można uzyskać tekst. Gdy nie można uzyskać żadnego tekstu, lepiej pozbyć się danych dotyczących tej osoby badanej i zacząć od początku.

Opisywana metoda jest jak na razie jasna i niekontrowersyjna. Doszliśmy do mało porywającego wniosku, że jesteśmy w stanie przetworzyć nagranie w tekst w sposób naukowy. Poświęciliśmy wiele czasu na upewnienie się co do tego, gdyż kolejny krok daje nam szansę empirycznego badania świadomości, jednocześnie stawiając przed nami liczne przeszkody oraz sprawiając trudności. Musimy wyjść poza tekst; musimy zinterpretować go jako zapis *aktów mowy*; nie zwykłej wymowy czy deklamacji, ale jako zapewnień, pytań, odpowiedzi, obietnic, komentarzy, próśb o wyjaśnienie, myślenia na głos czy uwag kierowanych do samego siebie.

Tego rodzaju interpretacja wymaga od nas przyjęcia czegoś, co nazywam *nastawieniem intencjonalnym* (Dennett 1971, 1978a, 1987a): musimy traktować wydającego z siebie głos osobnika jako podmiot działający, racjonalny, który ma swoje przekonania i pragnienia oraz inne stany umysłowe, które cechują się *intencjonalnością* czy „dotyczeniem" (*aboutness*), a których czyny można wyjaśnić (lub przewidzieć) na podstawie treści tychże stanów. Dlatego też tworzone dźwięki powinny być interpretowane jako coś, co osoba badana *chciała powiedzieć*, na przykład jako *sądy*, które chciała z różnych przyczyn *uznać*. Okazuje się, że już wcześniej opieraliśmy się na takim założeniu, kiedy oczyszczaliśmy tekst. (Zastanawialiśmy się: Dlaczego ktokolwiek *chciałby powiedzieć* „z dóry na gół"?)

Pomimo potencjalnych niebezpieczeństw pojawiających się w momencie przyjęcia nastawienia intencjonalnego w stosunku do tych zachowań werbalnych, są one ceną, jaką musimy zapłacić za zdobycie dojścia do szeregu wiarygodnych truizmów, które wykorzystujemy przy projektowaniu eksperymentów. Na przykład czasami ludzie chcą mówić pewne rzeczy nie dlatego, że w nie wierzą, ale dlatego, iż są przekonani, że chce je usłyszeć widownia. Zwykle warto zrobić oczywiste rzeczy, aby zmniejszyć prawdopodobieństwo pojawienia się lub wcielenia w życie tego pragnienia: mówimy podmiotom, że chcemy usłyszeć ich przekonania, i staramy się, aby nie dowiedzieli się, jakich przekonań oczekujemy. Innymi słowy, robimy wszystko, co w naszej mocy, aby znaleźli się w sytuacji, w której z pomocą wszczepionych im pragnień (pragnienia współpracy, bycia wynagrodzonym, bycia dobrym podmiotem) nie mieli lepszego wyjścia, niż po prostu próbować wyrazić swoje rzeczywiste przekonania.

Kolejne zastosowanie nastawienia intencjonalnego w stosunku do naszych osób badanych jest wymagane, jeśli mamy korzystać z tak użytecznych pomocy, jak naciskanie przycisku. Jest to zwykle sposób wykonania jakiegoś konwencjonalnie ustalonego aktu mowy, jak na przykład *uznania*, że dwa widziane obrazki ukazują mi się *teraz* nałożone na siebie, lub odpowiedź, że *tak*, mój pospieszny, błyskawiczny osąd (gdyż powiedziano mi, że szybkość jest tu najistotniejsza) jest taki, iż słowo, które właśnie słyszę, było na liście, którą odtwarzano chwilę temu. Dla wielu celów eksperymentalnych będziemy więc chcieli zinterpretować znaczenie owych wciśnięć przycisku i włączyć je do tekstu. To, jaki akt mowy będzie wynikiem konkretnego naciśnięcia przycisku, zależy od intencjonalnej interpretacji interakcji między osobą badaną a eksperymentatorem, która pojawiła się w czasie przygotowywania jej do eksperymentu. (Nie zawsze naciskanie przycisku jest aktem mowy; czasem może to być na przykład udawane strzelanie lub udawane sterowanie rakietą).

Kiedy pojawiają się wątpliwości, czy osoba badana powiedziała to, co chciała, czy zrozumiała problem lub czy rozumie znaczenie użytych słów, możemy poprosić o wyjaśnienie. Zwykle można w ten sposób rozwiać wątpliwości. W sytuacji idealnej efektem tych posunięć jest usunięcie wszelkich możliwych źródeł dwuznaczności i niepewności z sytuacji eksperymentalnej, tak aby bezkonkurencyjna okazała się tylko *jedna* intencjonalna interpretacja tekstu (łącznie z naciskaniem przycisków). Jest ona traktowana jako szczere, autentyczne wyrażenie przekonań i opinii przez *pojedynczą osobę badaną*[26]. Jak jednak zobaczymy, pojawią się momenty, w których to założenie może być problematyczne – zwłaszcza jeśli nasz badany wykazuje jakąś patologię. Jak na przykład powinniśmy odnosić się do ewidentnie szczerej skargi na ślepotę w przypadkach tak zwanej ślepoty histerycznej czy do niewątpliwie szczerego zaprzeczenia ślepocie u osób niewidomych z anosognozją (zaprzeczanie ślepocie lub zespół Antona)? Tym zjawiskom przyjrzymy się bliżej w kolejnych rozdziałach, bo jeśli mamy zrozumieć, czego doświadczają ci ludzie, nie wystarczy nam prosty wywiad.

4. Światy fikcyjne i heterofenomenologiczne

Poza problemami wynikającymi z odosobnionych przypadków wydaje się, że istnieje problem bardziej ogólny. Czy sama praktyka interpretowania zachowania werbalnego w podany sposób nie zakłada świadomości osoby badanej i czy wobec tego nie powracamy do pytania o zombi? Wyobraźmy sobie, że masz kontakt z „mówiącym" komputerem, i załóżmy, że udało ci się zinterpretować dane jako akty mowy wyrażające poglądy i opinie dotyczące domniemanych stanów świadomych. Sam fakt, że istnieje jedna spójna interpretacja ciągu zachowań, nie zakłada, że interpretacja ta jest *prawdziwa*; osoba badana mogła jedynie zachowywać się, *jak gdyby* była świadoma; ryzykujemy bycie oszukanymi przez zombi, które nie posiadają żadnego

życia wewnętrznego. Ten sposób interpretacji nie pozwoliłby ci *potwierdzić*, że komputer był czegokolwiek świadom. Nie możemy mieć pewności, że obserwowane przez nas akty mowy wyrażają prawdziwe przekonania dotyczące faktycznych przeżyć; być może wyrażają jedynie pozorne przekonania dotyczące *nieistniejących przeżyć*. Mimo to sam fakt, że znaleźliśmy choćby jedną stabilną interpretację zachowania jakiejś jednostki jako aktu mowy, zawsze będzie wart uwagi. Jeśli ktokolwiek odnalazł intersubiektywny, jednorodny sposób interpretacji poruszających się gałęzi drzewa na wietrze jako „komentarzy" wyrażanych przez „pogodę" na temat aktualnych wydarzeń politycznych, znalazł coś wspaniałego, co wymaga wyjaśnienia, nawet jeśli miałoby się okazać genialnym wynalazkiem stworzonym przez jakiegoś inżyniera-żartownisia.

Na szczęście istnieje analogia, która pomoże nam *opisać* tego rodzaju fakty bez konieczności ich *wyjaśniania*: możemy porównać zadanie heterofenomenologa dotyczące zachowania osoby badanej do zadania czytelnika, który interpretuje fikcję. O niektórych tekstach, takich jak powieści czy nowele, wiemy – lub przypuszczamy – że są fikcją, ale nie przeszkadza nam to w ich interpretacji. W pewnym sensie interpretacja jest w tych przypadkach prostsza, ponieważ nie uwzględnia, bądź odsuwa w czasie, trudne pytania dotyczące szczerości, prawdy czy odniesienia.

Spójrzmy na pewne niekontrowersyjne fakty dotyczące semantyki fikcji (Walton 1973; Lewis 1978; Howell 1979). Powieść opowiada historię, lecz historię nieprawdziwą, pomijając zbiegi okoliczności. Mimo wiedzy o nieprawdziwości tej historii mówimy o tym, co jest w historii *prawdziwe*. „Możemy faktycznie powiedzieć, że Sherlock Holmes mieszkał na Baker Street, oraz to, że lubił się popisywać swoimi zdolnościami intelektualnymi. Nie możemy powiedzieć, że poświęcał się dla rodziny lub że blisko współpracował z policją" (Lewis 1987, s. 37). W historii prawdziwe jest o wiele więcej niż tylko to, co bezpośrednio powiedziano w tekście. Faktem jest, że w Londynie z czasów Holmesa nie istnieją samoloty (choć nie jest to wyrażone dosłownie ani nie możemy tego w logiczny sposób wywnioskować z tekstu) oraz że istnieją stroiciele fortepianów (chociaż – o ile dobrze pamiętam – żaden nie jest w tekście wspomniany; i ponownie nie jest to fakt, który możemy racjonalnie z tekstu wydedukować). Poza kwestiami, które są prawdziwe lub nie w historii, istnieje spory obszar nieokreśloności: Holmes i Watson wsiedli do pociągu jadącego do Aldershot ze stacji Waterloo o godzinie 11.10, lecz nie jest ani prawdą, ani fałszem, że była to środa (*Garbus*).

Istnieją pewne subtelne problemy filozoficzne dotyczące tego, jak (ściśle) wyrazić rzeczy, o których chcemy mówić wprost, gdy rozmawiamy o fikcji, nie będziemy jednak się nimi zajmować. Być może niektórzy są głęboko zaniepokojeni metafizycznym statusem bohaterów czy przedmiotów, ale ja nie. Jestem beztroskim optymistą i nie uważam, aby istniały jakieś głębokie problemy filozoficzne dotyczące tego, jak powinniśmy ontologicznie odnosić się do produktów fikcji; fikcja to *fikcja*; Sherlock Holmes *nie istnieje*. Zostawiając więc na marginesie zawiłości oraz błyskotliwe, techniczne propozycje radzenia sobie z nią, chciałbym zwrócić waszą uwagę na prosty fakt: interpretacja fikcji jest niezaprzeczalnie wykonalna i prowadzi do pewnych niekontrowersyjnych rezultatów. Po pierwsze, rozumienie historii, czyli na przykład odkrywanie „świata Sherlocka Holmesa", nie jest jałowe czy niewykonalne; czytelnik może sporo dowiedzieć się o powieści, tekście i jego celu, o autorze, a nawet o prawdziwym świecie, poznając *świat przedstawiony* w powieści. Po drugie, jeśli uważnie zidentyfikujemy i wyeliminujemy sądy wyrażające gusty i preferencje (np. „Watson to nudny pedant"), możemy zebrać wiele niekwestionowanie obiektywnych faktów o świecie przedstawionym. Wszyscy odbiorcy zgadzają się, że Holmes był mądrzejszy od Watsona; obiektywność polega na nudnej oczywistości.

Po trzecie – co jest ulgą dla uczniów – wiedza o świecie przedstawionym w powieści

może być niezależna od znajomości tekstu. Byłbym prawdopodobnie w stanie napisać pracę semestralną o *Pani Bovary*, choć nigdy nie czytałem powieści – nawet angielskiego tłumaczenia. Widziałem serial BBC, więc znam fabułę. Wiem, co się dzieje w tamtym świecie. Sedno jest następujące: fakty dotyczące świata fikcji są faktami na *poziomie semantycznym* owej fikcji; nie są zależne od faktów syntaktycznych tekstu (jeśli zakładamy, że fikcja to tekst). Możemy porównać przedstawienie teatralne czy film *West Side Story* ze sztuką Szekspira *Romeo i Julia*; opisując podobieństwa i różnice dotyczące tego, co wydarza się w tych światach, widzimy podobieństwa między dziełami sztuki, których nie sposób opisać w kategoriach odpowiadających syntaktycznym czy tekstualnym (a już na pewno nie fizycznym) opisom konkretnych przypadków z fikcji. Fakt, że w każdym z tych światów jest para kochanków, którzy należą do różnych frakcji, nie dotyczy słownictwa, struktury zdania, długości (w słowach lub klatkach filmowych) czy też wielkości, kształtu ani wagi żadnego z konkretnych fizycznych przykładów tych dzieł.

Ogólnie rzecz biorąc, możemy opisać, co zostało przedstawione w dziele sztuki (na przykład w *Pani Bovary*) niezależnie od opisywania, jak tego dokonano. (Oczywiście zwykle nie próbujemy tego od siebie odseparować i mieszamy komentarz do świata przedstawionego z komentarzem dotyczącym sposobu przedstawienia go przez autora, ale ta separacja jest możliwa). Możemy sobie nawet wyobrazić, że wiemy tyle o przedstawionym świecie, że jesteśmy w stanie zidentyfikować autora fikcji, nie znając tekstu ani jakiegokolwiek jego tłumaczenia. Dowiadując się pośrednio, co się dzieje w fikcji, możemy być gotowi na stwierdzenie: tylko Wodehouse mógł wymyślić tak niedorzeczny przypadek. Wydaje nam się, że potrafimy zidentyfikować rodzaje wydarzeń i okoliczności (a nie tylko rodzaje *opisów* wydarzeń i okoliczności) jako kafkowskie i jesteśmy gotowi sklasyfikować bohaterów jako typowo szekspirowskich. Wiele z tych wiarygodnych przekonań jest bez wątpienia mylnych (jak mogłyby pokazać pomysłowe eksperymenty), ale nie wszystkie. Wspominam je po to tylko, aby zilustrować, jak wiele informacji jesteśmy w stanie zgromadzić z *tego, co zaprezentowano*, pomimo posiadania skąpej wiedzy dotyczącej tego, *jak reprezentacja* została osiągnięta.

Zastosujmy teraz analogię do problemu, z jakim styka się eksperymentator, który chce zinterpretować tekst stworzony przez osoby badane bez odnoszenia się do pytania o to, czy osoby badane to zombi, komputery, kłamcy czy osoby zdezorientowane. Rozpatrzmy zalety przyjęcia taktyki interpretacji tych tekstów jako swego rodzaju fikcji, oczywiście nie literackich, ale jako generatorów *fikcji teoretycznych* (co koniec końców może się okazać prawdą). Czytelnik powieści pozwala tekstowi *tworzyć* (fikcyjny) świat, wyznaczony z definicji przez tekst w sposób wyczerpująco ekstrapolowany i prowadzący nieograniczenie dalej; nasz eksperymentator, heterofenomenolog, pozwala osobie badanej na to, aby jej tekst *stanowił heterofenomenologiczny świat osoby badanej*, świat wyznaczony z definicji przez tekst (zinterpretowany) i rozszerzający się nieokreślenie dalej. Pozwala to heterofenomenologowi odsunąć w czasie skomplikowane problemy dotyczące tego, jakie mogłyby być relacje pomiędzy tym (fikcyjnym) światem a światem prawdziwym. Pozwala to teoretykom zgodzić się w szczegółach co do tego, czym jest heterofenomenologiczny świat osoby badanej, jednocześnie nie zgadzając się na jedno ujęcie tego, jak światy heterofenomenologiczne odnoszą się do zdarzeń mózgowych (czy też duszy). Heterofenomenologiczny świat osoby badanej będzie stabilnym, intersubiektywnie potwierdzalnym teoretycznym postulatem, mającym ten sam status metafizyczny co na przykład Londyn Sherlocka Holmesa czy świat według Garpa.

Jak w fikcji, tak i tu liczy się to, co powie autor (lub rzekomy autor). Ściślej rzecz biorąc, to, co powie rzekomy autor, dostarcza nam tekstu, który po zinterpretowaniu według reguł właśnie przeze mnie opisanych ustala, jaki pewien świat *jest*. Nie pytamy, jak Conan Doyle

dowiedział się o kolorze fotela Holmesa, i nie bierzemy pod uwagę możliwości, że odgadł go źle; poprawiamy za to literówki lub wyszukujemy najlepsze, najspójniejsze odczytanie tekstu, jakie możemy znaleźć. Nie pytamy również, skąd osoby badane (pozorne osoby badane) wiedzą to, o czym mówią, i nawet w najmniejszym stopniu nie próbujemy (na chwilę obecną) twierdzić, że się mylą; bierzemy je za (zinterpretowane) słowo. Warto też zwrócić uwagę, że powieści często rozpoczynają się klauzulą, zapewniającą o przypadkowości ewentualnych zbieżności z prawdziwymi ludźmi, ale taktyka pozwalania tekstowi na stwarzanie świata nie musi być ograniczona do dzieł literackich, które autorzy pisali jako fikcję; możemy opisać królową Wiktorię pewnego biografa lub świat Henry'ego Kissingera, beztrosko lekceważąc przypuszczalne intencje autora, aby mówić prawdę i odnosić się, nieprzypadkowo, do prawdziwych ludzi.

5. Dyskretny urok antropologa

Traktowanie ludzi jako generatorów fikcji (teoretycznych) nie jest czymś powszechnym. Przyznawanie władzy konstytutywnej ich twórczości może być postrzegane jako protekcjonalne, prześmiewcze i nieszczere. Widać to wyraźnie w trochę innym zastosowaniu taktyki heterofenomenologicznej przez antropologów. Wyjaśni nam to następujący przykład. Wyobraź sobie, że antropolodzy odkryli plemię wierzące w boga lasu, Feenomana, o którym dotychczas nie słyszeli. Dowiadując się o nim, antropolodzy stanęli przed fundamentalnym wyborem: mogą przyjąć religię tubylców i całym sercem uwierzyć we wspaniałe czyny Feenomana lub badać ten kult z agnostycznego punktu widzenia. Rozważmy drugą możliwość. Mimo że nie wierzą w Feenomana, antropolodzy postanawiają najlepiej, jak potrafią, zbadać i usystematyzować nowo poznaną religię. Notują opisy Feenomana dyktowane im przez wierzących. Starają się odnaleźć spójną postać, lecz nie zawsze im się to udaje (niektórzy twierdzą, że Feenoman ma oczy niebieskie, inni, że jest on – lub ona – brązowooki). Próbują wyjaśnić i wyeliminować te nieporozumienia, identyfikując i ignorując wśród osób badanych szukających poklasku przygłupów, rozważając różne interpretacje z pomocą miejscowych informatorów, a może nawet prowadząc mediacyjne dysputy. Stopniowo wyłania się logiczna konstrukcja: Feenoman to leśny bóg posiadający szereg cech i nawyków oraz własną biografię. Owi agnostyczni naukowcy (nazywający siebie feenomenologami) opisali, uporządkowali i skatalogowali część świata stworzonego przez wierzenia tubylców oraz (jeśli naprawdę przyłożyli się do pracy) wykonali *ostateczny* opis Feenomana. Wierzenia tubylców (nazwijmy ich feenomanistami) są autorytatywne (jest w końcu *ich* bogiem), ale tylko dlatego, że Feenoman jest traktowany *jedynie* jako „przedmiot intencjonalny", zwykła fikcja dla niewierzących, więc w zupełności jako *istota* wierzeń (prawdziwych bądź nie) feenomanistów. Owe wierzenia mogą być ze sobą sprzeczne, więc Feenoman, jako logiczna struktura, może mieć przypisane sprzeczne cechy – ale nie jest to problemem w oczach feenomenologów, ponieważ jest on dla nich *jedynie* strukturą. Feenomenolodzy starają się przedstawić najlepszy logiczny konstrukt, jaki potrafią zbudować, ale nie mają obowiązku rozwiązać wszystkich sprzeczności. Są przygotowani na odkrycie nierozwiązywalnych nieporozumień wśród wierzących.

Feenomaniści oczywiście nie widzą tego w ten sposób – z definicji, ponieważ dla nich Feenoman nie jest jedynie przedmiotem intencjonalnym, lecz kimś tak prawdziwym, jak ty czy ja. Ich stosunek do swojego autorytetu dotyczącego cech Feenomana jest – czy też powinien być – trochę bardziej skomplikowany. Z jednej strony wierzą w to, że *wiedzą* o Feenomanie wszystko – w końcu są feenomanistami, więc kto może wiedzieć lepiej niż oni? Lecz jeśli nie uważają się za kogoś nieomylnego niczym papież, dopuszczają ewentualne błędy w szczegółach.

Prawdopodobnie mogliby zostać poinformowani o prawdziwej naturze Feenomana. Na przykład sam Feenoman mógłby wyjaśnić im parę szczegółów. Mogliby więc być trochę zaniepokojeni pozbawioną wrażeń łatwowiernością (tak to rozumieją) zadających pytania feenomenologów, którzy niemal zawsze wierzą im na słowo, nigdy nie kwestionując tego, co słyszą, nie wątpiąc, jedynie z szacunkiem pytając, jak można wyjaśnić pojawiające się niejednoznaczności i konflikty.

Metoda heterofenomenologiczna ani nie kwestionuje zapewnień osób badanych, ani nie przyjmuje ich jako całkowicie prawdziwych, ale stara się zachować konstruktywną i współczującą neutralność, mając nadzieję na stworzenie *ostatecznego* opisu świata według osób badanych. Którakolwiek z osób badanych, zaniepokojona przyznaniem takiej konstytutywnej władzy, mogłaby zaprotestować: „Ale *naprawdę*! Rzeczy, które wam opisuję, są *zupełnie prawdziwe* i mają właśnie takie cechy, o jakich wam mówię!". Szczera odpowiedź heterofenomenologa mogłaby polegać na kiwnięciu głową i zapewnieniu osoby badanej, że nikt nie wątpi w jej prawdomówność. Jednak wierzący zwykle chcą więcej – chcą, aby publiczność zaufała ich zapewnieniom, a jeśli to się nie udaje, chcą wiedzieć, kiedy im nie wierzymy – dlatego najbardziej dyplomatycznym podejściem dla heterofenomenologów, bez względu na to, czy to antropolodzy, czy eksperymentatorzy badający świadomość w laboratorium, jest unikanie zwracania uwagi na swoją oficjalną neutralność.

Owo odchylenie od zwykłych relacji interpersonalnych to cena, jaką musimy zapłacić za neutralność wymaganą od nas przez naukę o świadomości. Oficjalnie musimy uważać, czy nasze osoby badane to przypadkiem nie kłamcy, zombi czy papugi w ludzkim stroju, ale nie musimy narażać ich na zmartwienie, głośno ten fakt oznajmiając. Poza tym owa taktyka neutralności jest jedynie czasowym przystankiem na drodze do stworzenia i potwierdzenia empirycznej teorii, która w zasadzie mogłaby potwierdzić słuszność twierdzeń osób badanych.

6. Odkrywanie tego, o czym ktoś tak naprawdę mówi

Czym byłoby potwierdzenie wiary osoby badanej w jej własną fenomenologię? Możliwości widzimy lepiej, odnosząc się do naszych analogii. Rozważmy, w jaki sposób moglibyśmy potwierdzić, że jakaś „powieść" jest w rzeczywistości prawdziwą (lub w większości prawdziwą) biografią. Moglibyśmy na początku zadać pytanie: na jakiej prawdziwej osobie z kręgu swoich znajomych wzorował się autor? Czy to w rzeczywistości ukryta matka autora? Jakie prawdziwe wydarzenia z dzieciństwa autora zostały przetworzone w tym fragmencie? Co *tak naprawdę* autor próbuje powiedzieć? Zapytanie o to autora może nie być najlepszą drogą do odpowiedzi na te pytania, gdyż może on tego po prostu nie wiedzieć. Czasem można brać pod uwagę, iż autor został zmuszony, nie zdając sobie z tego sprawy, do alegorycznego czy metaforycznego wyrażenia siebie. Jedyne źródła, które autor był w stanie wykorzystać, nie pozwoliły – z jakiegoś powodu – na bezpośrednie, faktyczne, niemetaforyczne przedstawienie zdarzeń, o których chciał opowiedzieć; stworzona przez niego historia to kompromis. Może on być drastycznie zreinterpretowany (nawet przy protestach udręczonego autora), ukazując prawdziwą opowieść o faktycznych ludziach i wydarzeniach. Można często stwierdzić, że nie jest przypadkiem, iż pewna fikcyjna postać posiada dane cechy, co pozwala nam zreinterpretować tekst, który prezentuje tę postać w sposób taki, że jej cechy mogą się odnosić – w prawdziwym, niefikcyjnym kontekście – do cech i czynów prawdziwej osoby. Sportretowanie fikcyjnej Molly jako zdziry może być prawidłowo odebrane jako zniesławienie prawdziwej Polly, gdyż wszystko, co dotyczy Molly, *tak naprawdę* dotyczy Polly. Protesty autora co do takiej interpretacji mogą, słusznie lub nie, przekonać nas, iż to oszczerstwo nie jest świadome bądź zamierzone, ale od

dawna powiadają nam Freud i inni, że autorzy, tak jak my sami, często nie mają pojęcia o głębszych źródłach swoich intencji. Jeżeli istnieje nieświadome oszczerstwo, musi również istnieć nieświadomy dla niego kontekst.

Wracając do naszej poprzedniej analogii, wyobraźmy sobie, co by się stało, gdyby antropolog potwierdził, że rzeczywiście istnieje niebieskooka postać o imieniu Feenoman, która leczy chorych i przeskakuje z drzewa na drzewo jak Tarzan. To nie bóg i nie potrafi latać ani być w dwóch miejscach naraz, wciąż jednak stanowi niewątpliwie prawdziwe źródło większości relacji, legend i wierzeń feenomanistów. Naturalnie byłoby to powodem wielu bolesnych rozczarowań wśród wiernych, z których spora grupa zrewidowałaby swoją wiarę, choć inni pozostaliby przy ortodoksyjnej wersji, nawet jeśli oznaczałoby to utrzymanie „prawdziwego" Feenomana (z jego nadnaturalnymi zdolnościami) i jednoczesne uznanie istnienia jego odpowiednika z krwi i kości. Można zrozumieć opór ortodoksów przed myślą, że mylili się co do Feenomana. Jeśli kandydat antropologów na rzeczywistego Feenomana z doktryny wierzących nie posiadałby niemal identycznych cech i czynów jak Feenoman z legendy, nie pozwoliliby na stwierdzenie żadnego takiego odkrycia. (Porównaj: „Odkryłem, że Święty Mikołaj jest prawdziwy. W rzeczywistości jest wysokim, chudym skrzypkiem z Miami i nazywa się Fred Dudley; nienawidzi dzieci i nigdy nie kupuje prezentów").

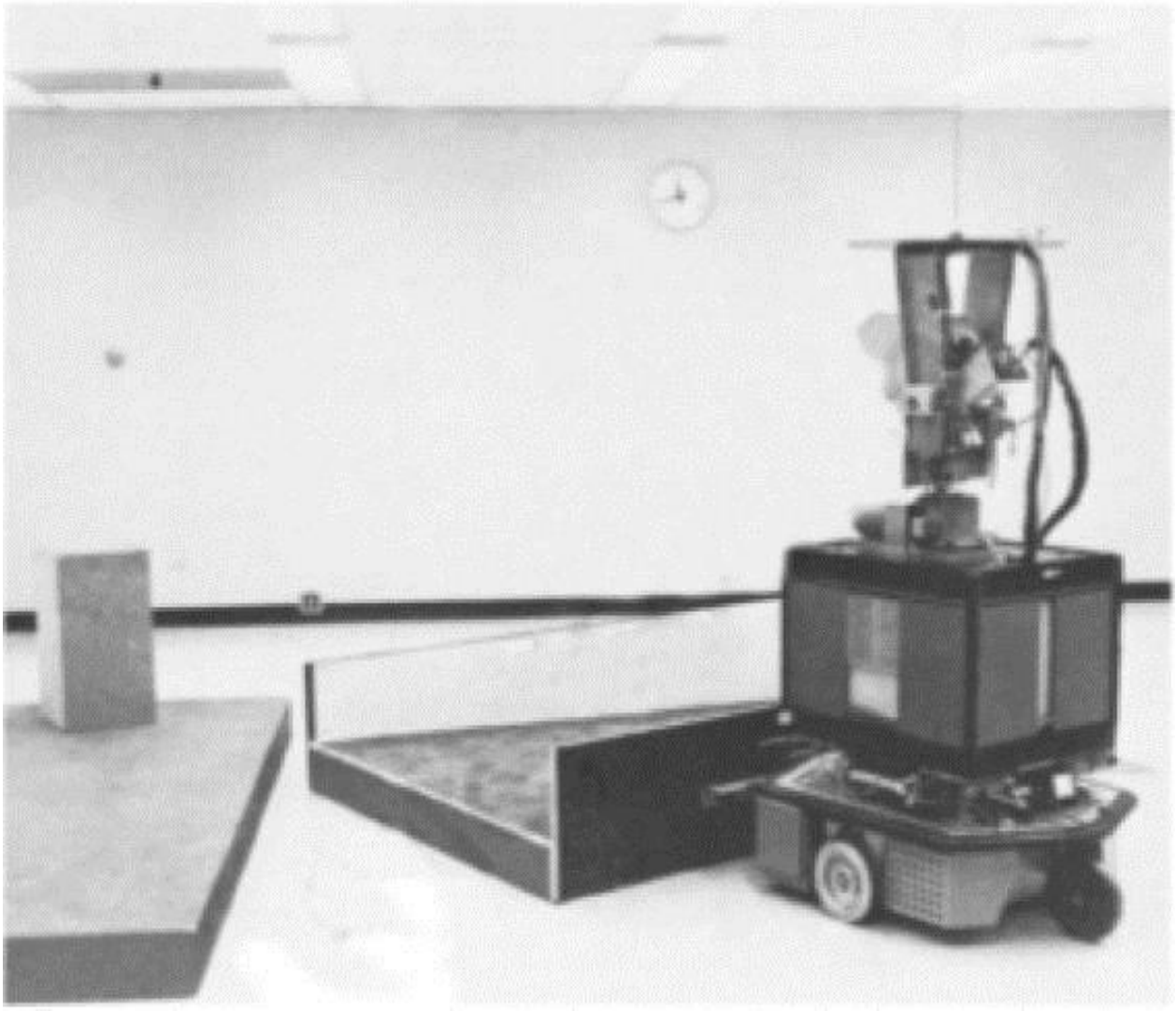
Moja sugestia jest zatem taka, że gdybyśmy znaleźli w mózgach ludzi prawdziwe zdarzenia, które miałyby *wystarczającą ilość* „definiujących" cech składników wypełniających ich heterofenomenologiczne światy, moglibyśmy racjonalnie stwierdzić, że odkryliśmy to, o czym *naprawdę* myślą – nawet jeśli z początku ludzie opieraliby się takim identyfikacjom. Gdybyśmy natomiast odkryli, że prawdziwe wydarzenia są tylko w minimalny sposób podobne do składników heterofenomenologicznych, moglibyśmy racjonalnie stwierdzić, że ludzie po prostu mylili się co do wierzeń, które wyrażali, pomimo swojej szczerości. Ktoś zawsze mógłby powiedzieć – na przykład zawzięci feenomaniści – że prawdziwe składniki fenomenologiczne *towarzyszyły* zdarzeniom i nie były z nimi identyczne, natomiast inną kwestią jest to, czy takie twierdzenie byłoby dla kogokolwiek przekonujące.

Tak jak antropolodzy, i my możemy pozostać neutralni w badaniach. Ta neutralność może się wydawać bezcelowa – czy nie jest niemal nie do wyobrażenia, żeby naukowcy mogli odkryć zjawiska neurofizjologiczne, które *po prostu* byłyby składnikami uznawanymi przez osoby badane w ich heterofenomenologiach? Zdarzenia mózgowe wydają się zbyt różne od elementów fenomenologicznych, aby mogły być prawdziwymi odpowiednikami przekonań, które wyrażamy w naszych introspekcyjnych relacjach. (Jak widzieliśmy w rozdziale 1, substancja umysłowa zdaje się potrzebna, aby tworzyć z niej na przykład fioletowe krowy). Przypuszczam, iż większość ludzi cały czas uznaje perspektywy takiej identyfikacji za absolutnie niewyobrażalne, lecz zamiast przyznać, że jest to w związku z tym niemożliwe, chcę spróbować trochę rozruszać naszą wyobraźnię, opowiadając jeszcze jedną historyjkę. Zbliży ona nas powoli do szczególnie skomplikowanego składnika fenomenologicznego, jakim jest *obraz umysłowy*. Ten składnik jest prawdziwy, choć historyjka jest nieco uproszczona i przerysowana.

7. Obrazy umysłowe Shakeya

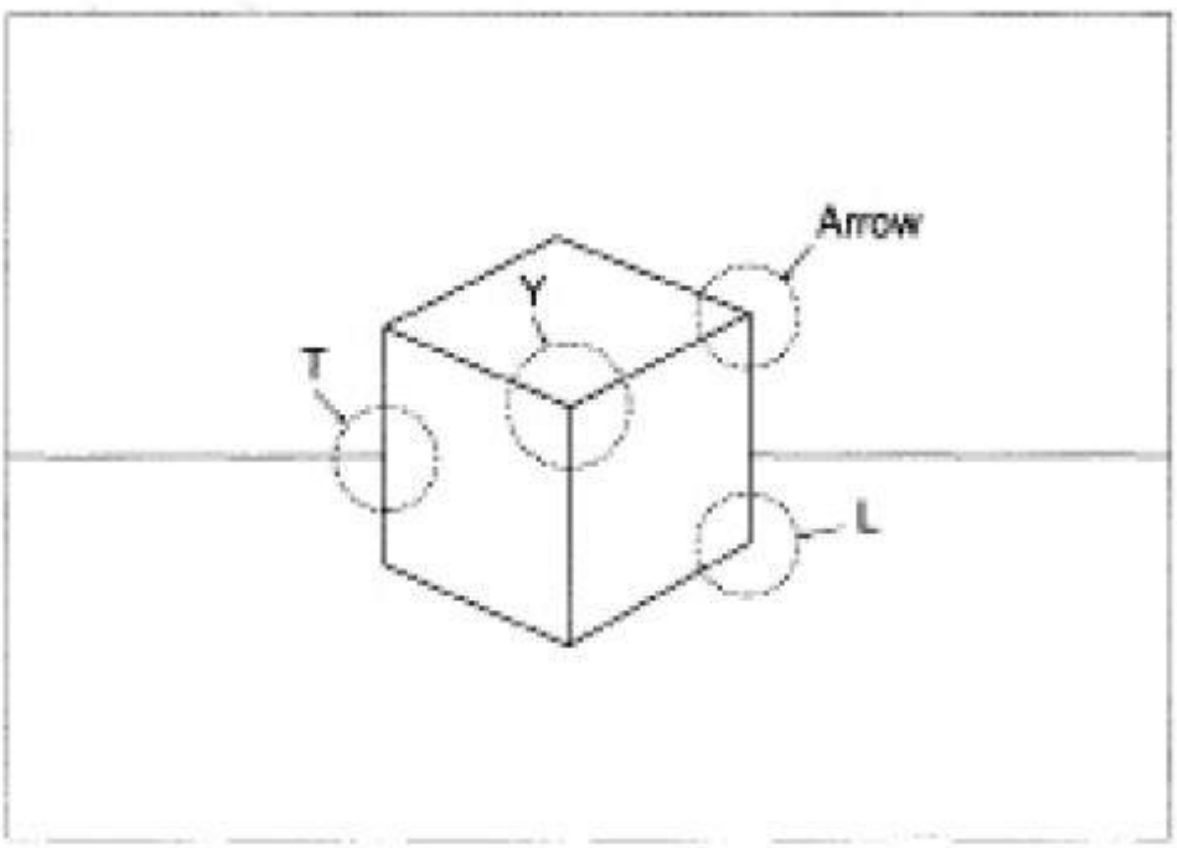
W krótkiej historii robotów Shakey, skonstruowany w Stanford Research Institute w Menlo Park w Kalifornii w późnych latach sześćdziesiątych XX wieku przez Nilsa Nilssona, Bertrama Raphaela i ich współpracowników, zasługuje na status legendarnego nie dlatego, że wykonał coś niezwykle poprawnie ani że był szczególnie udaną symulacją jakiejś ludzkiej cechy psychicznej, ale dlatego, iż w przedziwny sposób otworzył pewne drogi myśli, jednocześnie

zamykając inne (Raphael 1976; Nilsson 1984). Był rodzajem robota, którego mógł podziwiać filozof, swego rodzaju argumentem na kółkach.



Ryc. 4.1

Shakey był pudełkiem na kółkach z telewizyjnym okiem i zamiast nosić swój mózg ze sobą, był z nim połączony (był to w tamtych czasach duży komputer stacjonarny) poprzez radio. Shakey zamieszkiwał kilka pokoi, w których jedynymi przedmiotami było kilka pudeł, piramid, ramp i platform, a wszystkie były pomalowane i oświetlone, aby Shakeyowi łatwiej było „widzieć". Z Shakeyem można było się komunikować, wpisując wiadomości do terminalu podczepionego pod jego komputerowy mózg i stosując ściśle ograniczone słownictwo w języku angielskim. Polecenie „ZEPCHNIJ PUDŁO Z PLATFORMY" sprawiało, że Shakey szukał pudła, lokalizował rampę, którą podsuwał sobie w odpowiednie miejsce, a następnie wjeżdżał po niej na platformę, z której spychał pudło.



Ryc. 4.2

Jak Shakey to robił? Czy w jego wnętrzu siedział może karzełek, który obserwował ekran telewizyjny i naciskał kontrolne przyciski? Taki pojedynczy, bystry homunkulus byłby jednym ze sposobów – kłamliwych – na sukces. Innym sposobem byłoby umieszczenie ludzkiego kontrolera poza Shakeyem, w pilocie radiowym. Byłoby to rozwiązanie kartezjańskie, w którym rolę szyszynki odgrywałby nadajnik/odbiornik zlokalizowany w Shakeyu, a rolę niefizycznych poleceń z duszy przejęłyby wcale nienadprzyrodzone sygnały radiowe. Bezwartościowość tych „rozwiązań" jest oczywista; jakie natomiast mogłoby być rozwiązanie wartościowe? Z początku może się wydawać niepojmowalne – a przynajmniej wysoce skomplikowane – ale to właśnie z takimi przeszkodami w naszej wyobraźni musimy się skonfrontować i je pokonać. Okazuje się o wiele łatwiejsze, niż mogło ci się z początku wydawać, w jaki sposób Shakey wykonywał te zadania bez pomocy *homo ex machina*.

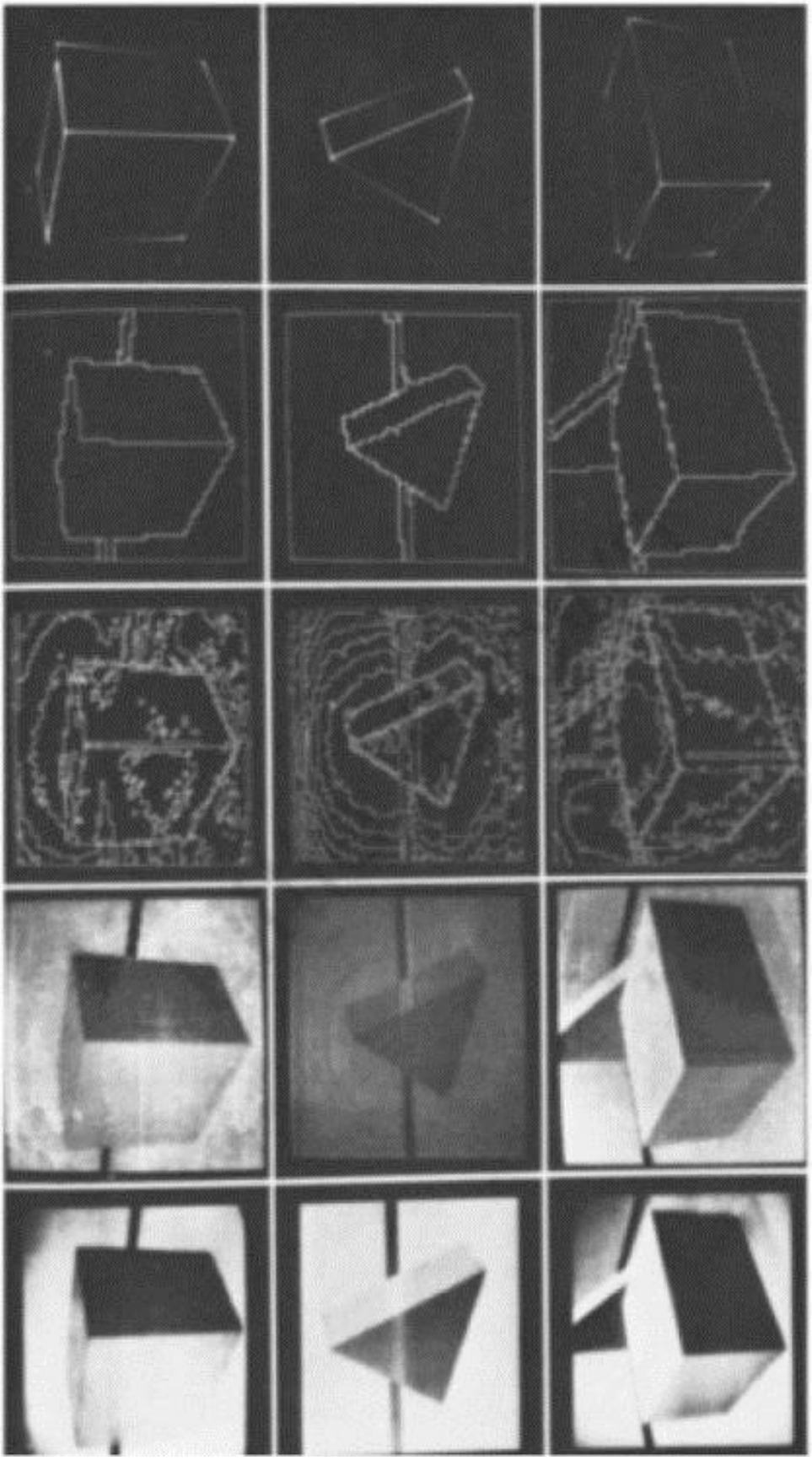


Рис. 4.3

W jaki sposób Shakey odróżniał pudła od piramid okiem telewizyjnym? Zarys odpowiedzi był widoczny dla obserwatorów, którzy przyglądali się procesowi na monitorze komputerowym. Jedno ujęcie widoczne na ziarnistym ekranie, przedstawiające na przykład pudełko, pojawiało się na ekranie; następnie obraz był oczyszczany, prostowany i wyostrzany na różne sposoby, po czym brzegi pudełka były wyrysowywane na biało i cały obraz zamieniał się w obraz liniowy (Ryc. 4.3).

Następnie Shakey analizował liniowy obraz: każdy wierzchołek był identyfikowany jako L, T, X, strzałka lub Y. Jeśli został odkryty wierzchołek Y, przedmiot musiał być pudełkiem, nie piramidą; z żadnej perspektywy piramida nie mogła ukazywać wierzchołka Y.

Jest to pewnego rodzaju uproszczenie, ale pokazuje ono generalne zasady funkcjonowania robota; Shakey korzystał z programu „semantyki linii", który posługiwał się ogólnymi zasadami, aby określić kategorię obiektu, którego obraz pojawiał się na ekranie. Obserwując go, obecni mogli nagle poczuć się zdezorientowani, gdy w końcu docierało do nich, że dzieje się coś dziwnego: oni obserwowali proces transformacji obrazu, ale Shakey nie spoglądał na żaden inny monitor, na którym obraz ulegał zmianie i był analizowany. W jego konstrukcji nie było żadnych innych monitorów i dlatego monitor, który obserwowali, mógł zostać wyłączony lub odłączony od prądu i nie miałoby to żadnego wpływu na proces analizy percepcyjnej Shakeya. Czy ten monitor był jakimś oszukaństwem? Po co został stworzony? Jedynie dla obserwatorów. Czy zatem wydarzenia obserwowane na monitorze miały coś wspólnego z tym, co działo się wewnątrz Shakeya?

Monitor był dla obserwatorów, ale *pomysł* na niego miał również służyć projektantom Shakeya. Warto sobie uświadomić niemal niewyobrażalne zadanie, które przed nimi stało: w jaki sposób można za pomocą informacji z kamery telewizyjnej stworzyć rzetelny sposób identyfikacji pudełek? Z nieograniczonej liczby możliwych ujęć, które kamera była w stanie przesłać do komputera, na niewielu z nich znalazły się pudełka; każda klatka składa się jedynie z sieci czarnych i białych pikseli, komórek włączonych i wyłączonych, zer i jedynek. Jak można napisać program, który znajdowałby wszystkie klatki reprezentujące pudełka? Załóżmy, trochę przesadzając, że siatkówka kamery to sieć 10 000 pikseli, 100 na 100. Wówczas każde ujęcie byłoby jednym z 10 000 możliwych sekwencji zer i jedynek. Jakie układy zer i jedynek ukazywałyby się nam w obecności pudełka?

Zacznijmy od umieszczenia tych zer i jedynek na tablicy, która odzwierciedla obraz kamery w przestrzeni, jak sieć pikseli widocznych na ekranie. Ponumerujmy piksele w każdym rzędzie od lewej do prawej, jak słowa na kartce (inaczej niż w telewizji komercyjnej, w której skan jest ukośny). Następnie zauważmy, że ciemne regiony składają się głównie z zer, a jasne z jedynek. Co więcej, *pionowa* granica między rejonem jasnym z lewej strony a rejonem ciemnym z prawej może być łatwo opisana za pomocą ciągów zer i jedynek: ciąg prawie samych jedynek aż do piksela *n*, a następnie ciąg głównie zer, po którym dokładnie 100 cyfr później (w kolejnej linii) następuje kolejna sekwencja głównie jedynek do piksela $n + 100$, po czym ciąg głównie zer, i tak dalej, co setkę.

Ryc. 4.4

Program, który wyszukiwałby taką okresowość w ciągu cyfr przychodzących z kamery, byłby w stanie zlokalizować takie pionowe granice. Gdy taka granica zostałaby już znaleziona, mogłaby zostać zamieniona w wyraźną białą linię przez rozsądne podmienianie zer z jedynkami i *vice versa* tak, żeby coś na kształt 00011000 ukazywało się dokładnie co 100 pozycji w ciągu.

Ryc. 4.5

Poziomą granicę między jasnością i ciemnością można zauważyć równie łatwo: miejsce w ciągu, w którym przeważająca większość jedynek w 100, 200 czy 300 cyfrach, po których następuje 100, 200 czy 300 cyfr, wśród których większość to zera.

Ryc. 4.6

Ukośne granice są tylko trochę bardziej skomplikowane; program musi szukać przesunięć w sekwencjach. Kiedy już wszystkie granice zostaną zlokalizowane i wyrysowane na biało, rysunek linii jest kompletny, wykonuje się kolejny, bardziej wyrafinowany krok: „szablony" są „umieszczane" na fragmentach segmentów linii, aby zidentyfikować wierzchołki. Wówczas prostym zadaniem jest użycie programu semantyki linii w celu skategoryzowania obiektu na obrazie – w niektórych przypadkach może to być zadanie tak łatwe, jak znalezienie pojedynczego wierzchołka Y.

Kilka właściwości tego procesu jest dla nas ważnych. Po pierwsze, każdy z tych podprocesów jest „banalny" i mechaniczny. To znaczy, że żadna część komputera nie musi rozumieć, co robi i dlaczego, i nie ma żadnej zagadki w tym, jak każdy z kroków jest mechanicznie wykonywany. Mimo to sprytna organizacja tych banalnych, mechanicznych procesów tworzy urządzenie, które *zastępuje* obeznanego obserwatora. (Umieśćmy cały system wzroku w „czarnej skrzynce", której zadaniem jest „powiedzieć Shakeyowi to, co potrzebuje wiedzieć" o tym, co znajduje się przed nim, na podstawie informacji pochodzących z klatki telewizyjnej. Początkowo mogło nam się wydawać, że konieczne jest umieszczenie w czarnej skrzynce małego człowieczka, który obserwuje ekran. Teraz jednak widzimy, że ten homunkulus ma do wykonania banalną pracę i może być zastąpiony przez maszynę).

Gdy już wiemy, na czym polega ten proces, widzimy, że podczas gdy jest on *analogiczny* do procesu faktycznego obserwowania (oraz rysowania i wymazywania) białych i czarnych kropek na ekranie, faktyczne położenie w komputerze pojedynczych operacji zamiany zer na jedynki i *vice versa* nie ma znaczenia, dopóki liczby, które są tymczasowymi „adresami" indywidualnych cyfr, kodują informacje o sąsiedztwie pikseli. Załóżmy, że wyłączymy monitor. Wówczas, mimo że faktyczny dwuwymiarowy obraz w konkretnym miejscu komputera (swego rodzaju „wzór pobudzenia sprzętu") nie istnieje (lub nie musi istnieć), operacje te są homomorficzne (analogiczne) do zdarzeń oglądanych na monitorze. Owe zdarzenia były autentycznymi wyobrażeniami: dwuwymiarowa powierzchnia pobudzonych punktów luminescencyjnych, tworzących kształt o konkretnej wielkości, kolorze, lokalizacji i orientacji. Zatem w *pewnym* ścisłym sensie Shakey *nie odnajduje* pudełek przez serię transformacji obrazu; ostatni realny obraz w tym procesie to ten znajdujący się w receptywnym polu widzenia kamery. W innym ścisłym, choć metaforycznym sensie Shakey *odnajduje* pudełka przez serię transformacji obrazu – w ramach właśnie opisanego procesu, w którym jasno-ciemne granice są zmieniane w wyrysowane linie, a następnie kategoryzowane jako wierzchołki. Fakt, że ten ścisły

sens jest również metaforyczny, można wyjaśnić stwierdzeniem, że istnieje wiele cech, które spodziewamy się dostrzec w *prawdziwych* obrazach, a nieobecnych w obrazach Shakeya: nie mają koloru, wielkości, orientacji. (Moglibyśmy stworzyć zagadkę z takiego obrazu: myślę o obrazie, który nie jest ani mniejszy, ani większy od *Mony Lisy*, nie jest ani kolorowy, ani czarno-biały oraz nie jest skierowany w żadną stronę świata. Co to takiego?)

Proces, z którego korzystał Shakey, aby uzyskać informacje o obiektach ze światła znajdujących się w otoczeniu, nie przypominał właściwie w ogóle widzenia u człowieka ani prawdopodobnie żadnej innej istoty. Jednak na chwilę możemy to zignorować, aby dostrzec dość abstrakcyjną możliwość dotyczącą tego, jak obrazy umysłowe, o których opowiadają ludzie, mogą zostać odkryte w mózgu. Opis systemu widzenia Shakeya był uproszczony w celu unaocznienia podstawowych kwestii teoretycznych. Teraz posłużymy się fantastyką naukową, by pokazać coś jeszcze: wyobraźmy sobie połączenie Shakeya z inną znaną postacią o sztucznej inteligencji, SHRDLU Terry'ego Winograda (1972). SHRDLU przesuwał (wyimaginowane) klocki, po czym odpowiadał na pytania o to, co robi i dlaczego. Odpowiedzi SHRLDU były w dużej mierze zaprogramowane – przechowywane jako gotowe do użycia zdania oraz szablony zdań stworzone przez Winograda. Celem SHRDLU było abstrakcyjne badanie pewnych zadań przetwarzania informacji, z którymi mierzy się każdy rozmówca, a nie stworzenie realistycznego modelu ludzkiej mowy, co pozostaje w duchu naszego eksperymentu myślowego. (W rozdziale 8 przyjrzymy się bardziej realistycznym modelom produkcji mowy). Wymiana zdań z naszą nową wersją Shakeya, przeprojektowanego tak, aby mógł posługiwać się bardziej wyszukanymi czynnościami mowy, mogłaby wyglądać tak:

Dlaczego przesunąłeś rampę?

ŻEBYM MÓGŁ WJECHAĆ NA PLATFORMĘ.

A dlaczego chciałeś to zrobić?

ŻEBY ZEPCHNĄĆ PUDEŁKO.

A dlaczego chciałeś zepchnąć pudełko?

BO MI KAZAŁEŚ.

Następnie zapytalibyśmy Shakeya:

Jak odróżniasz pudełka od piramid?

Jaką odpowiedź na to pytanie moglibyśmy zaprogramować w Shakeyu? Przedstawiam trzy możliwości:

(1) Skanuję każdą sekwencję 10 000 zer i jedynek z mojej kamery, szukając pewnych wzorów sekwencji, jak na przykład... bla, bla, bla (bardzo długa odpowiedź, jeśli pozwolimy Shakeyowi na zagłębianie się w szczegóły).

(2) Znajduję granice między polem jasnym i ciemnym oraz rysuję w głowie białe linie wzdłuż nich; następnie patrzę na wierzchołki; jeśli na przykład znajdę wierzchołek Y, to wiem, że to pudełko.

(3) Nie wiem; niektóre rzeczy po prostu wyglądają jak pudełka. Po prostu je rozpoznaję. Dzięki *intuicji*.

Jaka powinna być poprawna odpowiedź Shakeya? Każda z nich jest na swój sposób prawdziwa; są opisami przetwarzania informacji na różnym poziomie. To, którą z odpowiedzi zaprogramujemy Shakeyowi, jest w dużej mierze kwestią decyzji, jak duży dostęp ma funkcja wyrażania się Shakeya (jego czarna skrzynka, czyli SHRDLU) do jego procesów percepcyjnych. Być może istnieją powody, aby nie miał on całkowitego dostępu (szczegółowego

i czasochłonnego) do pośrednich procesów analizy. Bez względu na to, jakimi możliwościami opisu samego siebie obdarzymy Shakeya, jest granica zaawansowania i szczegółowości jego wyrażalnej „wiedzy" na temat tego, co się w nim dzieje, co robi. Jeśli najlepsza odpowiedź, jaką potrafi dać, to (3), wówczas jest w takiej sytuacji, mówiąc o rozróżnianiu piramid od pudeł, w której jesteśmy my, gdy ktoś nas zapyta, w jaki sposób rozróżniamy słowo „szum" od słowa „sum"; nie wiemy, jak to robimy; jedno brzmi jak „szum", a drugie jak „sum" – to najlepsza odpowiedź, jaką potrafimy dać. Jeśli natomiast zaprojektujemy Shakeya tak, aby podał odpowiedź (2), nadal będą istniały pytania, na które nie będzie umiał odpowiedzieć, na przykład: „Jak rysujesz linię na swoich obrazach umysłowych?" lub „Jak rozpoznajesz, że jakaś krawędź to strzałka?".

Przyjmijmy, że programujemy Shakeya tak, aby miał dostęp typu (2) do swoich procesów analizy percepcyjnej; gdy go pytamy, jak to robi, opowiada nam o wykonywanym przez siebie transformowaniu obrazów. Bez jego wiedzy włączamy monitor. Czy mamy prawo powiedzieć mu, że wiemy lepiej? Tak naprawdę nie przetwarza obrazów, ale mu się tak wydaje? (*Mówi*, że tak robi, więc, stosując strategię heterofenomenologiczną, interpretujemy to jako wyraz jego przekonań). Gdyby był realistyczną symulacją osoby, mógłby nam odpowiedzieć, że nie mamy prawa mówić mu, co się dzieje w jego własnym umyśle! *On* wiedział, co robił, co *naprawdę* robił! Gdyby był bardziej wyrafinowany, mógłby stwierdzić, że to, co robił, mogłoby być jedynie alegorycznie opisywane jako przetwarzanie obrazu – chociaż właśnie taki opis przebiegu tego procesu wydawał mu się najtrafniejszy. W takim razie moglibyśmy mu powiedzieć, że jego metaforyczny sposób ujęcia tej kwestii był bardzo trafny.

Gdybyśmy jednak byli bardziej złośliwi, moglibyśmy zaprojektować Shakeya tak, aby zupełnie przypadkowo opisywał, co robi. Moglibyśmy zaprogramować go, aby opowiadał w sposób, który nijak nie odzwierciedla rzeczywistości („Wykorzystuję informacje z telewizora, aby prowadzić wewnętrzne dłuto, którym ciosam trójwymiarowy kształt z gliny mojej duszy. Wówczas, jeśli mój homunkulus może na nim usiąść, jest to pudełko; jeśli spada, to piramida"). Nie istniałaby żadna prawdziwa interpretacja jego opowieści; Shakey by po prostu *konfabulował* – wymyślałby historię, „nie zdając sobie z tego sprawy".

Ta możliwość w odniesieniu do nas pokazuje, dlaczego warto zadać sobie tyle trudu i traktować heterofenomenologię analogicznie do interpretacji fikcji. Jak już widzieliśmy, ludzie czasem po prostu mylą się co do tego, co robią i jak to robią. Nie *kłamią* w eksperymencie, ale konfabulują; zapełniają luki, zgadują, spekulują, mylą teoretyzowanie z obserwacją. Naprawdę niejasna pozostaje relacja między tym, co mówią, a tym, co sprawia, że to mówią – zarówno dla nas, heterofenomenologów, stojących *na zewnątrz*, jak i dla samych osób badanych. Nie mają one żadnej możliwości „widzenia" (choćby wewnętrznym okiem) procesów, które rządzą ich wypowiedziami, ale nie powstrzymuje ich to przed posiadaniem i wyrażaniem opinii, które głęboko w sobie czują.

Podsumowując, osoby badane są nieświadomymi twórcami fikcji, lecz mówienie, że są nieświadome, jest założeniem, że to, co wyrażają, może dokładnie zdawać sprawę *z tego, co im się wydaje*. Opowiadają, *czym dla nich jest* rozwiązywanie problemu, podejmowanie decyzji, rozpoznawanie obiektu. Jako że są (ewidentnie) szczere, zakładamy, że właśnie tak muszą te czynności odbierać, ale z tego wynika, że ich odczucia w najlepszym wypadku niejasno wskazują, co się w nich dzieje. Czasem owe nieświadome fikcje, które kreujemy my, osoby badane, mogą się okazać mimo wszystko prawdziwe, jeśli pozwolimy sobie na metaforyczne rozluźnienie, jak w przypadku odpowiedzi (2) Shakeya. Na przykład badania wyobrażeń umysłowych przeprowadzane przez psychologów poznawczych pokazują, że introspekcyjne twierdzenia dotyczące obrazów umysłowych, które sprawiają nam przyjemność (fioletowych

krów czy piramid), nie są zupełnie nieprawdziwe (Shepard i Cooper 1982; Kosslyn 1980; Kosslyn, Holtzman, Gazzaniga i Farah 1985). Przyjrzymy się temu szczegółowo w rozdziale 10 i zobaczymy, jak można interpretować introspekcyjne raporty dotyczące wyobrażeń, aby okazywały się prawdziwe. Jednak tak jak w przypadku ziemskiego Feenomana, który nie potrafił latać czy być w dwóch miejscach naraz, faktyczne zdarzenia odkryte w mózgu i *utożsamione* z obrazami umysłowymi nie będą miały owych cudownych właściwości, które nadały im osoby badane. „Obrazy" Shakeya świadczą o tym, że może istnieć coś, co w ogóle nie jest obrazem, mimo że jest opisywane jako obraz. Procesy mózgowe odpowiedzialne za wyobrażenia u ludzi prawdopodobnie nie są podobne do tego, co przebiega w Shakeyu, ale otworzyliśmy możliwości, które wcześniej były trudne do wyobrażenia.

8. Neutralność heterofenomenologii

Na początku tego rozdziału obiecałem opisać metodę, metodę heterofenomenologiczną, która byłaby neutralna w stosunku do debat o subiektywnych i obiektywnych podejściach do fenomenologii oraz o fizycznej i niefizycznej realności bytów fenomenologicznych. Przeanalizujmy tę metodę, aby stwierdzić, czy rzeczywiście taka jest.

Przede wszystkim: co z problemem zombi? Otóż heterofenomenologia sama w sobie nie może odróżnić zombi od prawdziwych, świadomych ludzi, nie twierdzi więc, że rozwiązała problem zombi, ani go nie pomija. *Ex hypothesi*, zombi zachowują się tak, jak normalni ludzie, a ponieważ heterofenomenologia jest sposobem interpretacji zachowania (w tym wewnętrznego zachowania mózgów itd.), stwierdzi, że światy Zoe i zombi-Zoe, jej nieświadomej bliźniaczki, są dokładnie takie same. Zombi mają świat heterofenomenologiczny, ale znaczy to tyle, że gdy teoretycy interpretują go, robią to tak samo, używając dokładnie tych samych środków co my, gdy interpretujemy naszych przyjaciół. Oczywiście, jak już powiedziano wcześniej, niektórzy z naszych przyjaciół mogą być zombi. (Trudno mi jest zachować powagę, ale ponieważ niektórzy bardzo poważni filozofowie podchodzą do tego problemu nie na żarty, czuję się w obowiązku postąpić podobnie).

Z pewnością nie ma nic złego czy nienaturalnego w przyznaniu zombi prawa do świata heterofenomenologicznego, gdyż to tak niewiele. To metafizyczny minimalizm heterofenomenologii. Metoda ta opisuje świat, heterofenomenologiczny świat osoby badanej, w którym znajdują się przeróżne *przedmioty* (przedmioty intencjonalne w żargonie filozoficznym) i w którym różne rzeczy wydarzają się tym przedmiotom. Jeśli ktoś zapyta: „Czym *są* te obiekty i *z czego* są zrobione?", odpowiedź *mogłaby* brzmieć: „Niczym!". Z czego zrobiony jest pan Pickwick? Z niczego. Pan Pickwick to fikcyjny obiekt, tak jak obiekty opisywane, nazywane, wspominane przez heterofenomenologów.

„Czy nie jest jednak wstydem przyznać, że jako teoretyk opowiadasz o istotach fikcyjnych – o rzeczach, które nie istnieją?" Wcale nie. Teoretycy literatury wykonują cenną, rzetelną robotę intelektualną, opisując istoty fikcyjne tak jak antropolodzy, którzy badają bogów czy czarownice w różnych kulturach. Robią to również fizycy, którzy zapytani, z czego jest zrobiony środek ciężkości, odpowiedzieliby: „Z niczego!". Tak jak środek ciężkości i równik, obiekty heterofenomenologiczne są *abstraktami*, nie *konkretami* (Dennett 1987a, 1991a/2008). Nie są czczymi fantazjami, ale żmudnie wypracowanymi fikcjami teoretycznymi. Poza tym, w przeciwieństwie do środka ciężkości, mamy otwartą drogę do tego, aby zamienić je na konkrety, jeśli postęp w naukach empirycznych nas do tego upoważni.

Są dwa sposoby badania potopu Noego: można założyć, że jest to po prostu mit, ale taki, który można doskonale zbadać, lub spytać, czy jest poparty dowodami jakiejś meteorologicznej

czy geologicznej katastrofy. Oba podejścia mogą być naukowe, ale pierwsze jest mniej oparte na domysłach. Drugi rodzaj spekulacji wymaga przeprowadzenia dokładnych badań pierwszego typu, aby dowiedzieć się, jakie można tam znaleźć podpowiedzi. Analogicznie, aby badać, jak (a nawet *czy*) obiekty fenomenologiczne to rzeczywiście zdarzenia mózgowe, najpierw należy heterofenomenologicznie skatalogować te obiekty. Ryzykuje się tu urażenie osób badanych (tak jak antropolodzy badający Feenomana ryzykują urażenie swoich informatorów), lecz tylko tak można uniknąć walki na „intuicje", które są uznawane za fenomenologię.

Co natomiast z obiekcją, iż heterofenomenologia, zaczynając od perspektywy trzecioosobowej, pozostawia *prawdziwe* kwestie świadomości z boku? Jak widzieliśmy, uważa tak Nagel oraz filozof John Searle, który wyraźnie ostrzega przed moim podejściem: „Pamiętajmy, że w tych dyskusjach zawsze należy wymagać punktu widzenia pierwszej osoby. Pierwszym krokiem do operacjonalistycznego kuglarstwa jest próba zrozumienia, skąd możemy *wiedzieć*, jak to jest dla innych" (Searle 1980, s. 451). To się jednak nie dzieje. Zwróćmy uwagę, że dla heterofenomenologii to *ty masz ostatnie słowo*. To ty redagujesz, poprawiasz, wypierasz się nieprzemyślanych zdań, a cokolwiek stwierdzisz – jeśli unikasz przy tym aroganckiego teoretyzowania na temat przyczyn czy metafizycznego statusu obiektów, których istnienie relacjonujesz – zostaje uznane za mające konstytutywny autorytet, i to ty rozstrzygasz, co dzieje się w twoim heterofenomenologicznym świecie. To ty jesteś pisarzem, więc wszystko, co powiesz, jest prawdą. Czego więcej można by chcieć?

Jeśli chcesz, abyśmy *uwierzyli* we wszystko, co mówisz o swojej fenomenologii, prosisz nie tylko o to, byśmy brali cię na poważnie, ale również, abyśmy uznali twoją papieską nieomylność, więc prosisz o zbyt wiele. *Nie jesteś* autorytetem dotyczącym tego, co dzieje się w tobie, lecz tylko w kwestii, co *wydaje ci się*, że dzieje się w tobie, i dajemy ci absolutną, dyktatorską władzę nad zdawaniem relacji z tego, co tobie się wydaje, z tego, *jak to jest być tobą*. Jeśli zaznaczysz, że niewyrażalne jest to, jak to właściwie się tobie wydaje, my, heterofenomenolodzy, również przyjmiemy to do wiadomości. Czy moglibyśmy mieć lepsze podstawy do tego, aby wierzyć, że nie potrafisz opisać czegoś, niż to, że (1) nie opisujesz tego oraz (2) przyznajesz, że nie potrafisz tego opisać? Oczywiście możesz kłamać, jednak my będziemy chcieli ci wierzyć. Jeśli powiesz: „Nie twierdzę, że *ja* nie potrafię tego opisać; twierdzę, że jest to nie do opisania!" – my, heterofenomenolodzy, odpowiemy, że nie jesteś w stanie opisać tego *teraz*, a ponieważ jesteś jedyną osobą na pozycji opisującego, w tym momencie jest to nie do opisania. Być może później będziesz w stanie to opisać, ale wówczas będzie to oczywiście coś innego, coś możliwego do opisania.

Gdy stwierdzam, że obiekty heterofenomenologiczne są fikcją teoretyczną, możesz poczuć chęć (okazuje się, że wielu ją czuje) sprzeciwienia się temu i powiedzenia:

To jest *właśnie* to, co odróżnia obiekty prawdziwej fenomenologii od obiektów heterofenomenologii. Moje *auto*fenomenologiczne obiekty nie są fikcją – są zupełnie *prawdziwe*, chociaż nie mam pojęcia, z czego są zrobione. Gdy szczerze ci mówię, że wyobrażam sobie fioletową krowę, nie wytwarzam po prostu potoku słów z tym związanych (jak Shakey), sprytnie kombinując, aby były podobne do jakichś ledwo analogicznych wydarzeń fizycznych w moim mózgu; świadomie i z premedytacją zdaję relację z istnienia czegoś, co *naprawdę tam jest*! Dla mnie nie jest to po prostu fikcja teoretyczna!

Uważnie zanalizujmy tę wypowiedź. Nie jest to tylko nieświadome tworzenie ciągu słów. W pewnym sensie *tworzysz* nieświadomie ciąg słów; nie masz pojęcia, jak to robisz ani co bierze udział w tworzeniu tego ciągu. Wciąż jednak twierdzisz, że nie robisz tego *po prostu*; wiesz, *dlaczego* to robisz; *rozumiesz* ten ciąg słów i nie żartujesz. Zgadzam się. Dlatego właśnie to, co mówisz, składa się na świat heterofenomenologiczny. Gdybyś tylko wyrzucał z siebie jakieś

przypadkowe słowa, prawdopodobieństwo istnienia słów, które można by tak zinterpretować, byłoby znikome. Z pewnością istnieje dobre wyjaśnienie tego, jak i dlaczego mówisz to, co mówisz, wyjaśnienie, które tłumaczy różnicę między powiedzeniem czegoś po prostu a powiedzeniem czegoś na serio, *ale ty tego wyjaśnienia jeszcze nie masz*. Przynajmniej nie w całości. (W rozdziale 8 rozwiniemy tę kwestię). Prawdopodobnie mówisz o czymś prawdziwym, przynajmniej przez większość czasu. Zobaczmy, czy możemy dowiedzieć się, cóż to jest.

Te zapewnienia dla wielu osób nie są wystarczające. Niektórzy po prostu nie chcą stosować się do takich zasad. Na przykład ludzie bardzo religijni czują się urażeni, gdy ktoś choćby delikatnie zasugeruje, że *może* istnieć inna prawdziwa religia. Nie traktują oni agnostycyzmu jako neutralności, lecz jako zniewagę, ponieważ jedną z zasad ich wiary jest to, że niewiara jest grzechem. Ci, którzy w to wierzą, mają do tego prawo, jak również prawo (jeśli można to tak nazwać) do urażonych uczuć, co przydarza im się, gdy spotykają agnostyków i sceptyków, ale jeśli nie potrafią zapanować nad niepokojem, który czują, gdy dowiadują się, że ktoś (jeszcze) nie wierzy w to, co mówią, wykluczają się z badań naukowych.

W tym rozdziale rozwinęliśmy *neutralną* metodę badania i opisywania fenomenologii. Zakłada ona pozyskiwanie i oczyszczanie *tekstów* (ewidentnie) mówiących osób oraz użycie tych tekstów do stworzenia fikcji teoretycznej, *heterofenomenologicznego świata* osoby badanej. Ten fikcyjny świat pełen jest obrazów, wydarzeń, dźwięków, zapachów, dobrych i złych przeczuć oraz uczuć, w których istnienie w swoim strumieniu świadomości osoba badana (pozornie) szczerze wierzy. W najszerszym znaczeniu jest to neutralne przedstawienie tego, co znaczy *być* tą osobą badaną – w jej własnych słowach, zinterpretowane najlepiej jak potrafimy.

Pozyskawszy taką heterofenomenologię, teoretycy mogą następnie poświęcić się zagadnieniu tego, co mogłoby wyjaśniać *istnienie* tej heterofenomenologii ze wszystkimi zawartymi w niej szczegółami. Heterofenomenologia istnieje – tak bezspornie, jak powieści czy inne twory fikcji. Ludzie niewątpliwie wierzą, że przydarzają im się obrazy umysłowe, bóle, przeżycia percepcyjne oraz cała reszta, i *ten* fakt – to, w co ludzie wierzą oraz co opisują, gdy wyrażają swoje przekonania – to zjawisko, które każda naukowa teoria umysłu musi brać pod uwagę. Organizujemy dane dotyczące tego zjawiska jako fikcję teoretyczną, „przedmioty intencjonalne" w światach heterofenomenologicznych. Pytanie, czy przedstawione w ten sposób obiekty, wydarzenia i stany mózgu – czy nawet duszy – istnieją w rzeczywistości, jest kwestią empiryczną. Jeśli odkryjemy odpowiednie obiekty rzeczywiste, możemy stwierdzić, iż są od dawna poszukiwanymi odpowiednikami pojęć osób badanych; jeśli nie, będziemy musieli wyjaśnić, dlaczego osobom badanym wydaje się, że te obiekty istnieją.

Nasze założenia metodologiczne zostały przedstawione, więc możemy zająć się empiryczną teorią samą w sobie. Zaczniemy od problemu określania w czasie i chronologicznego porządkowania obiektów w naszych strumieniach świadomości. W rozdziale 5 zaprezentuję pierwszy zarys teorii i pokażę, jak radzi on sobie z pewnymi łatwymi kwestiami. W rozdziale 6 zobaczymy, jak ta teoria pozwala nam reinterpretować pewne znacznie bardziej skomplikowane zjawiska, z którymi borykają się teoretycy. W rozdziałach 7–9 teoria nabierze konkretniejszego kształtu, aby zapobiec błędnym interpretacjom i zarzutom. Tam też lepiej ukażę jej mocne strony.

Część druga

Empiryczna teoria umysłu

Rozdział 5

Wielokrotne szkice kontra teatr kartezjański

1. Punkt widzenia obserwatora

Nie istnieje komórka ani grupa komórek w mózgu o takiej anatomicznej czy funkcjonalnej wyższości, żeby mogła być podporą lub środkiem ciężkości całego systemu.
William James, 1890

Kapitanowie motorówek żeglujący wzdłuż niełatwego wybrzeża sterują zwykle w kierunku jakiegoś charakterystycznego punktu. Odnajdują widoczną, ale daleką boję, znajdującą się z grubsza w kierunku, w którym płyną, upewniają się, że na drodze do niej nie ma żadnych ukrytych przeszkód i kierują się wprost na nią. Jednak co jakiś czas ich czujność zostaje na tyle uśpiona, że zapominają o zmianie kierunku przed boją i wpływają wprost na nią! Sukces polegający na osiągnięciu mniej ważnego celu, jakim jest trafienie w punkt, okazuje się tak krzepiący, że przyćmiewa o wiele ważniejszy cel, czyli omijanie tarapatów.

W każdym świadomym umyśle istnieje *punkt widzenia*. Jest to jedno z podstawowych założeń, jakie przyjmujemy na temat umysłów – lub świadomości. Świadomy umysł to obserwator, który odbiera ograniczony podzbiór ze wszystkich dostępnych informacji. Obserwator przyswaja dostępne informacje w konkretnym, (mniej więcej) ciągłym ciągu czasów i miejsc we wszechświecie. Dla celów praktycznych możemy uznać punkt widzenia konkretnego świadomego podmiotu za coś takiego: punkt poruszający się w czasoprzestrzeni. Weźmy standardowe diagramy stosowane w fizyce i kosmologii przedstawiające przesunięcie Dopplera, czyli ugięcie światła pod wpływem grawitacji.

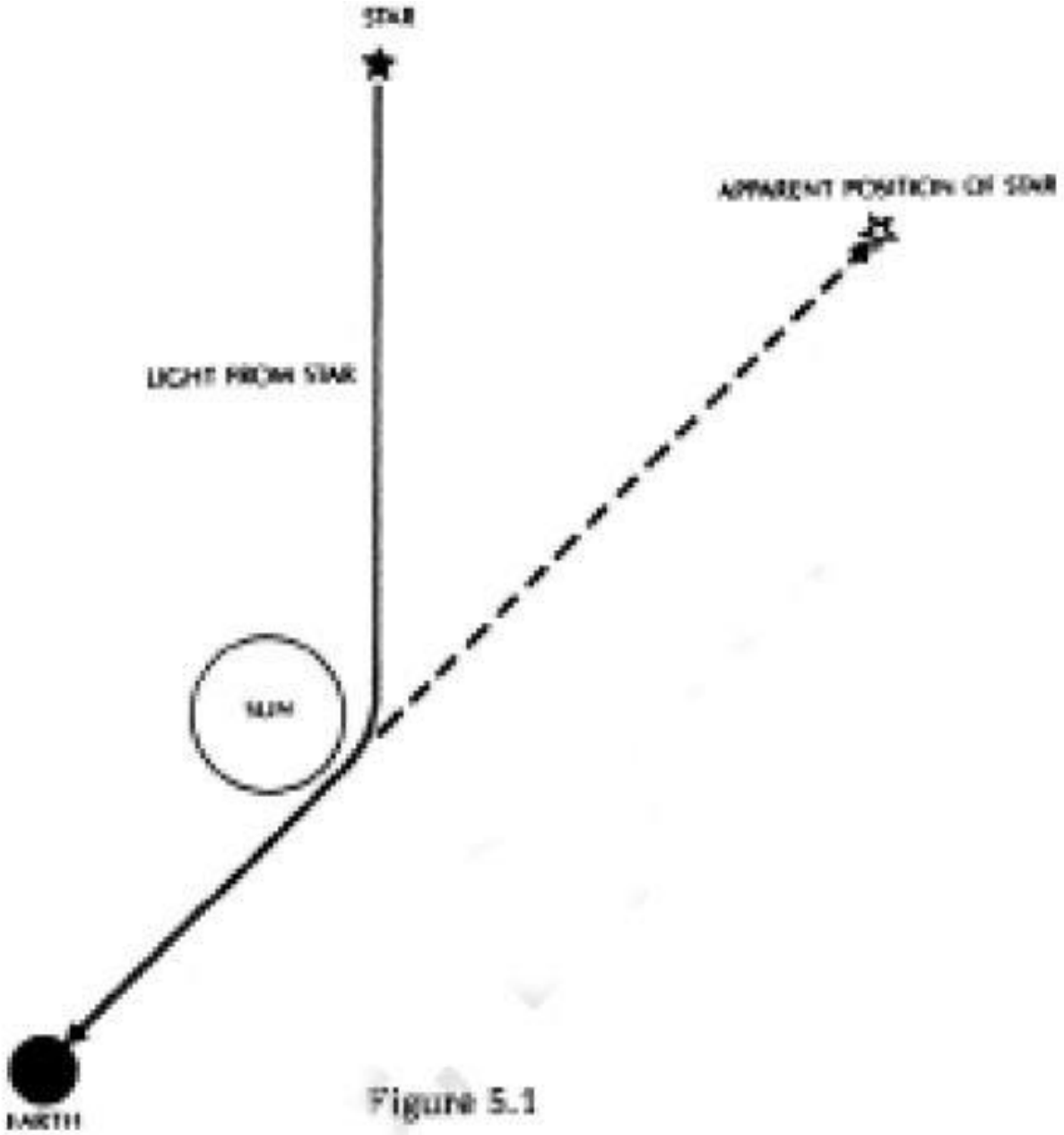


Figure 5.1

Ryc. 5.1

Obserwator na rycinie 5.1 znajduje się w punkcie na powierzchni Ziemi. Dla obserwatorów z różnych zakątków wszechświata sytuacja prezentowałaby się inaczej. Prostsze przykłady mogą być bardziej znajome. Różnicę między dotarciem do nas dźwięku i obrazu fajerwerków tłumaczymy różnymi prędkościami rozchodzenia się dźwięku i światła. Docierają *do obserwatora* (w konkretnym miejscu) w różnym czasie, choć ze źródła wydobyły się w tym samym momencie.

Cóż się jednak stanie, gdy skupimy się na obserwatorze i spróbujemy dokładnie zlokalizować jego punkt widzenia jako punkt *wewnątrz* jednostki? Proste założenia, które tak dobrze działają na dużą skalę, zaczynają się sypać[27]. Nie ma jednego punktu w mózgu, do którego docierają wszystkie informacje, a fakt ten niesie za sobą konsekwencje dalekie od

oczywistych – a nawet dosyć nieintuicyjne.

Będziemy zajmować się wydarzeniami zachodzącymi na względnie mikroskopijnej skali przestrzeni i czasu, jednak ważne, abyśmy rozumieli, jakie to będą wielkości. Wszystkie eksperymenty, o których będziemy wspominać, są związane z przedziałami czasu rzędu milisekund, czyli tysięcznych sekund. Pomocna będzie świadomość, jak długie (bądź krótkie) jest 100 ms czy 50 ms. Możesz wymówić cztery lub pięć sylab na sekundę, więc wymówienie jednej sylaby zabiera około 200 ms. Standardowe filmy wyświetlane są z prędkością 24 klatek na sekundę, więc klatki zamieniane są co 42 ms (dokładniej rzecz biorąc, każda klatka wyświetlana jest trzy razy podczas tych 42 ms, na czas 8,5 ms, z ciemną przerwą między nimi, trwającą 5,4 ms). Prędkość telewizji (w Stanach Zjednoczonych) to 30 klatek na sekundę, czyli jedna klatka na 33 ms (a dokładniej, każda klatka jest wpleciona w całość w dwóch miejscach, zazębiając się z klatką poprzednią). Twój kciuk w najlepszym wypadku może włączyć i wyłączyć stoper w 175 ms. Gdy uderzyć w palec młotkiem, szybkie włókna nerwowe (z osłonką mielinową) wysyłają wiadomość do mózgu w ciągu 20 ms; wolne włókna C bez osłonki mielinowej wysyłają znacznie później docierające do mózgu sygnały bólu – po około 500 ms – mimo tego samego dystansu do przebycia.

Poniżej znajduje się wykaz przybliżonych wartości milisekundowych różnych czynności:

powiedzenie „one, Mississippi”
1000 ms

włókno bez otoczki mielinowej, od czubka palca do mózgu
500 ms

wymówienie sylaby
200 ms

włączenie i wyłączenie stopera
175 ms

klatka filmowa
42 ms

klatka telewizyjna
33 ms

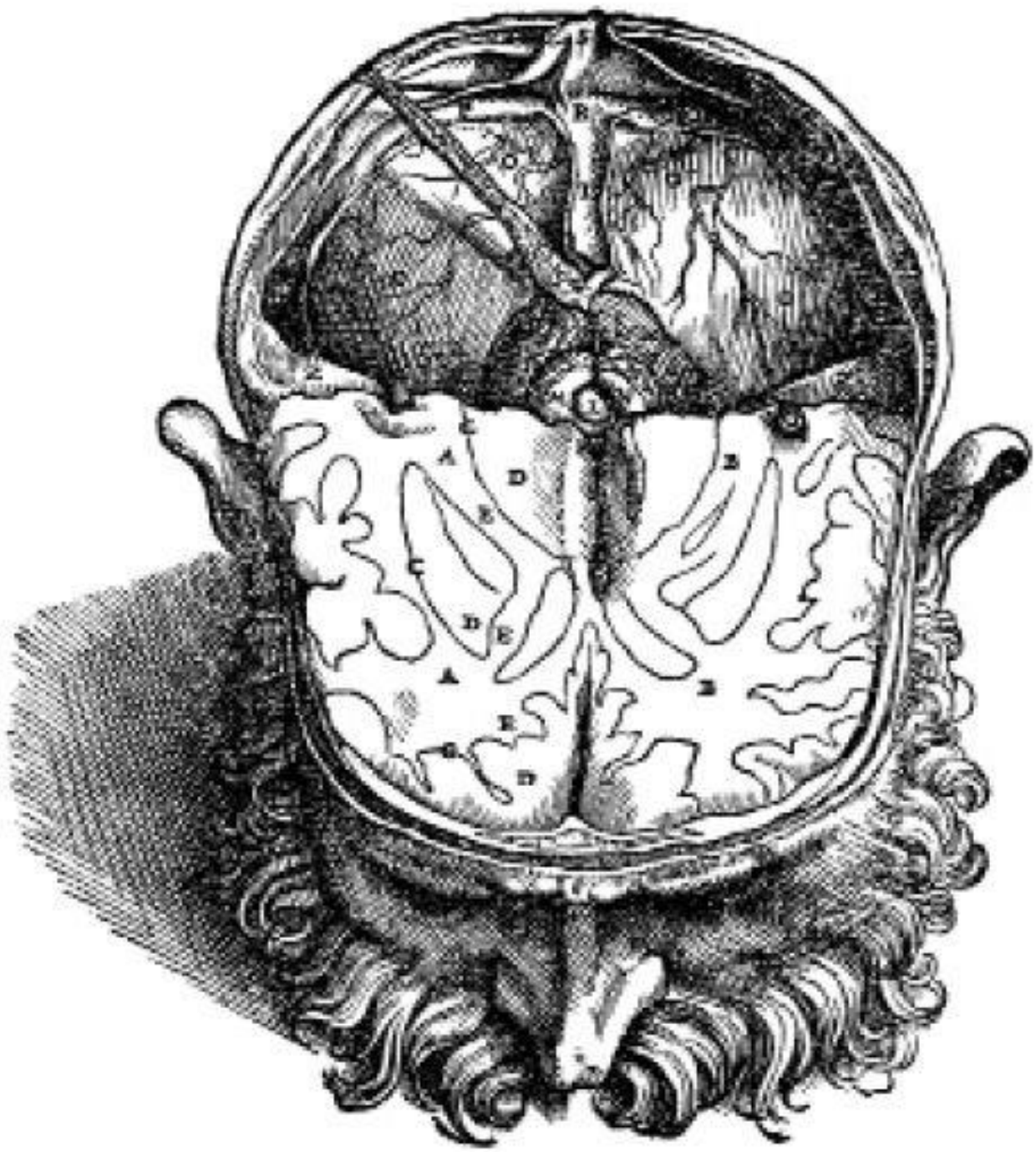
szybkie włókno (z otoczką mielinową), od czubka palca do mózgu
20 ms

czas trwania impulsu nerwowego
10 ms

czas trwania impulsu w komputerze cyfrowym
0,0001 ms

Kartezjusz jako pierwszy poważnie dociekał, co dostrzeżemy, przyglądając się wnętrzu ciała obserwatora, i wpadł na pomysł pozornie naturalny i pociągający. Ta idea przeniknęła myślenie o świadomości. Jak widzieliśmy w rozdziale 2, Kartezjusz stwierdził, że w naszym

mózgu znajduje się centrum – szyszynka będąca bramą do świadomego umysłu (Ryc. 2.1). Szyszynka to jedyna część mózgu, która jest zlokalizowana na pograniczu i nie ma lewej i prawej części. Na tej rycinie została zaznaczona przez wielkiego szesnastowiecznego anatoma Wesaliusza jako „L”. Jest mniejsza od ziarnka grochu, w zachwycającej izolacji od reszty opiera się na szypułce i jest połączona z układem nerwowym przez środek kresomózgowia. Jej funkcja była tajemnicza (do dziś nie wiadomo, czym dokładnie się zajmuje), więc Kartezjusz zaproponował dla niej następującą rolę: aby osoba była czegoś świadoma, informacje ze zmysłów muszą do niej docierać, a wówczas przeprowadza szczególne – a właściwie magiczne – oddziaływanie między materialnym mózgiem tej osoby a niematerialnym umysłem.



Ryc. 5.2

Według Kartezjusza nie wszystkie reakcje cielesne wymagają interwencji świadomego umysłu. Bardzo dobrze wiedział o istnieniu czegoś, co dziś nazywamy odruchami, i zakładał, że były one dokonywane na swego rodzaju mechaniczne skróty, które omijały przystanek „szyszynka” i w ten sposób przebiegały nieświadomie.

Ryc. 5.3

Mylił się co do detali: twierdził, że ogień poruszał skórę, co powodowało pociągnięcie za maleńką nitkę, a to z kolei otwierało pory w komorze (F), która wypuszczała „tchnienie żywotne" przez rurkę nadmuchującą mięśnie, co powodowało wycofanie stopy (Descartes 1662/1989). Poza tym był to jednak dobry pomysł. Nie można tego samego powiedzieć o jego wizji roli szyszynki jako kołowrotka świadomości (można by ją nazwać kartezjańskim wąskim gardłem). Ten pomysł, dualizm kartezjański, jest beznadziejnie błędny, jak widzieliśmy w rozdziale 2. Materializm takiego czy innego rodzaju jest obecnie powszechną opinią, niemalże

jednogłośnie przyjmowaną, ale nawet najbardziej wyrafinowani współcześni materialiści często zapominają, że gdy już pozbędziemy się upiornej kartezjańskiej *rei cogitantis*, nie trzeba szukać bramy ani jakiegokolwiek funkcyjnego centrum mózgu. Szyszynka nie tylko nie jest faksem łączącym z duszą, nie jest też gabinetem owalnym mózgu, i nie jest nim żadna inna część mózgu. Mózg to centrala, miejsce, w którym znajduje się ostateczny obserwator, lecz nie ma żadnego powodu, aby przypuszczać, że mózg posiada jakąś głębszą centralę, jakieś wewnętrzne sanktuarium, którego istnienie jest warunkiem koniecznym świadomego doświadczenia. Krótko mówiąc, w mózgu nie ma żadnego obserwatora[28].

Światło porusza się o wiele szybciej niż dźwięk, o czym przypominają nam na przykład fajerwerki, ale wiemy teraz, że mózg potrzebuje więcej czasu na przetworzenie bodźca wzrokowego niż słuchowego. Jak zauważył neuronaukowiec Ernst Pöppel (1985, 1989), dzięki tym równoważącym się różnicom „horyzont symultaniczności" to *około* 10 metrów: światło i dźwięk, które opuszczają ten sam punkt w odległości około 10 metrów od organów zmysłu obserwatora, wytwarzają neuronowe odpowiedzi, które są „centralnie dostępne" w tym samym momencie. Czy możemy to doprecyzować? Problemem nie jest jedynie zmierzenie odległości od źródła zewnętrznego do organów zmysłu, prędkość rozchodzenia się fal w różnych ośrodkach czy uwzględnienie indywidualnych różnic. Bardziej fundamentalną kwestią jest decyzja, co uznać za „linię mety" w mózgu. Pöppel uzyskał swój rezultat, porównując miary behawioralne: średni czas reakcji (wciskanie przycisku) na bodźce dźwiękowe i wizualne. Różnica oscyluje między 30 ms a 40 ms i jest to czas, jaki potrzebuje dźwięk na pokonanie 10 metrów (czas, jaki potrzebuje światło na pokonanie 10 metrów, jest zdecydowanie mniejszy od zera). Pöppel odwołał się do skrajnej linii mety – to zachowanie zewnętrzne – jednak nasza naturalna intuicja jest taka, że *przeżycie* światła czy dźwięku zachodzi *między* momentem, w którym wibracje docierają do naszych organów zmysłu, a momentem, w którym wciskamy przycisk, potwierdzając to przeżycie. Intuicja podpowiada nam również, że wydarza się to *centralnie*, gdzieś w mózgu, na pobudzonych ścieżkach między organem zmysłu a palcem. Wydaje się, że gdybyśmy potrafili powiedzieć *dokładnie* gdzie, moglibyśmy dokładnie stwierdzić, kiedy nastąpiło przeżycie. I *vice versa*: gdybyśmy potrafili powiedzieć, kiedy nastąpiło przeżycie, moglibyśmy stwierdzić, gdzie w mózgu zlokalizowane jest świadome przeżycie.

Ideę takiego centralnego miejsca w mózgu nazwiemy *materializmem kartezjańskim*, gdyż jest to pogląd, do którego dochodzi się po odrzuceniu dualizmu, ale z jednoczesnym zachowaniem wyobrażenia centralnego (choć materialnego) teatru, gdzie „wszystko się spotyka". Szyszynka jest jednym z potencjalnych urzeczywistnień teatru kartezjańskiego, lecz sugerowano również inne miejsca – przedni zakręt obręczy, układ siatkowaty, różne obszary płata czołowego. Materializm kartezjański zakłada, że istnieje ważna meta czy granica gdzieś w mózgu, będąca miejscem, gdzie kolejność przybycia jest równa kolejności „obecności" w przeżyciu, ponieważ to, co się tam zdarza, jest tym, co się przeżywa. Być może dziś nikt otwarcie nie głosi materializmu kartezjańskiego. Wielu teoretyków przekonywałoby nas, że całkowicie odrzucili ten wyraźnie błędny pomysł. Jak jednak zobaczymy, przekonujące wyobrażenie teatru kartezjańskiego wciąż powraca i nas nawiedza – zarówno amatorów, jak i naukowców – nawet po potępieniu i wyegzorcyzmowaniu duchowego dualizmu.

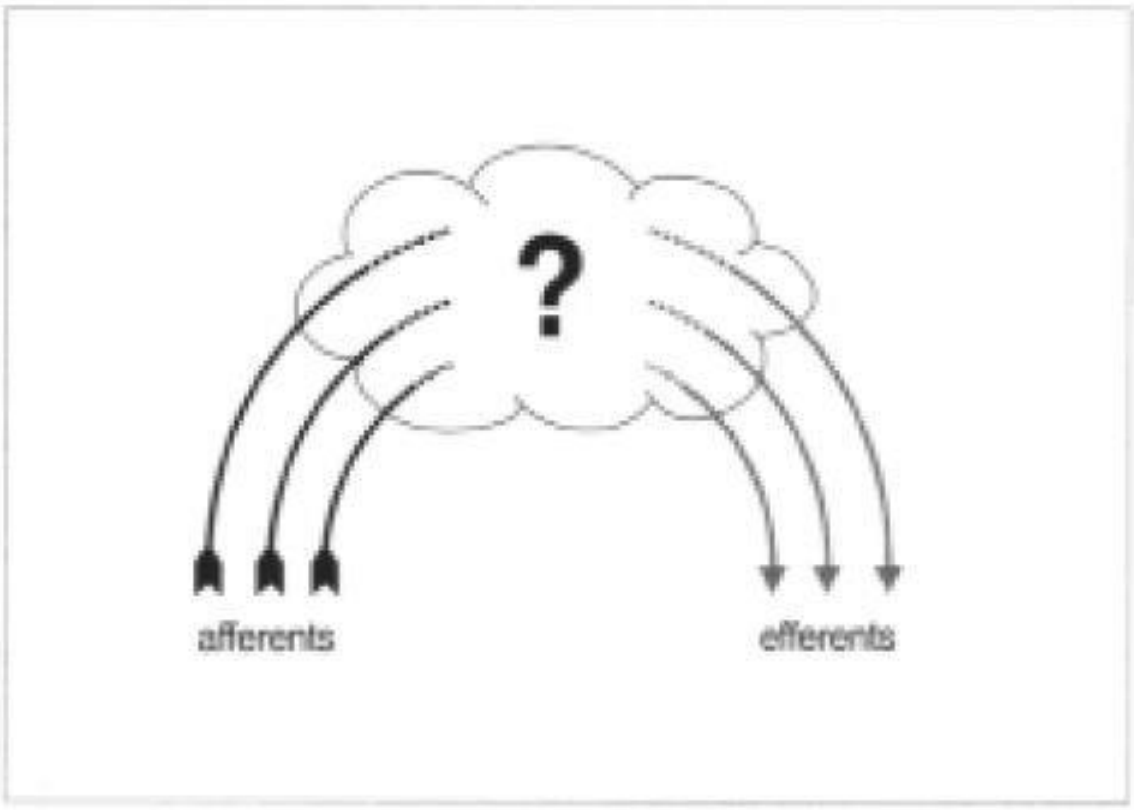
Teatr kartezjański jest metaforycznym obrazem lokalizacji świadomego przeżycia w mózgu. Początkowo wydaje się niewinną ekstrapolacją znanego i niezaprzeczalnego faktu, że *dla codziennych, makroskopijnych przedziałów czasowych* rzeczywiście możemy podzielić zdarzenia na dwie kategorie: „jeszcze niezaobserwowane" oraz „już zaobserwowane". Osiągamy to, lokując obserwatora w danym punkcie i zaznaczając ruchy nośników informacji w stosunku do tego punktu. Gdy jednak próbujemy użyć tej metody, aby wyjaśnić zjawisko przebiegające

w bardzo krótkich przedziałach czasu, napotykamy *logiczną* trudność: jeśli „punkt" widzenia obserwatora musi zostać rozłożony w mózgu o raczej sporej wielkości, subiektywne poczucie kolejności czy jednoczesności dla samego obserwatora musi być wyznaczone przez coś więcej niż sama „kolejność przybycia", gdyż kolejność ta nie jest ostatecznie zdefiniowana, dopóki nie zostanie ustalone miejsce tego przybycia. Jeśli A dotrze szybciej do jednej mety niż B, ale B dotrze przed A do innej, to który rezultat ustala subiektywną kolejność w świadomości? (Zob. także Minsky 1985, s. 61). Pöppel mówi o momentach, w których wzrok i słuch stają się „centralnie dostępne" w mózgu, jednak który punkt lub punkty „centralnej dostępności" miałyby się „liczyć" jako wyznaczniki *świadomej* kolejności i dlaczego właśnie one? Próbując odpowiedzieć na to pytanie, będziemy zmuszeni do porzucenia teatru kartezjańskiego i zastąpienia go nowym modelem.

Idea niezwykłego centrum w mózgu jest najtrwalszym złym pomysłem w dociekaniach natury świadomości. Jak zobaczymy, wciąż powraca ona w różnych przebraniach i jest przyjmowana z wielu pozornie przekonujących powodów. Przede wszystkim introspekcyjnie uznajemy własną „jedność świadomości", która sprawia wrażenie, jakby pozwalała odróżnić „tutaj" i „tam". Naiwna granica pomiędzy „mną" a „światem zewnętrznym" to skóra (i soczewki moich oczu). Im więcej wiemy, jak zdarzenia w naszych własnych ciałach mogą być dla „nas" niedostępne, tym bardziej wdziera się w nas świat zewnętrzny. „Tutaj" mogę próbować podnieść rękę, ale „tam" ręka „zdrętwiała" lub jest sparaliżowana i się nie porusza; naruszono linie łączności między *mną* a nerwową maszynerią sterującą ramieniem. Gdyby mój nerw optyczny został poważnie naruszony, nie spodziewałbym się, że nadal będę widzieć, choć moje oczy są zdrowe; przeżycie wzrokowe jest najwyraźniej czymś, co wydarza się głębiej niż w oku, gdzieś między tym organem a moim głosem, gdy opowiadam, co widzę.

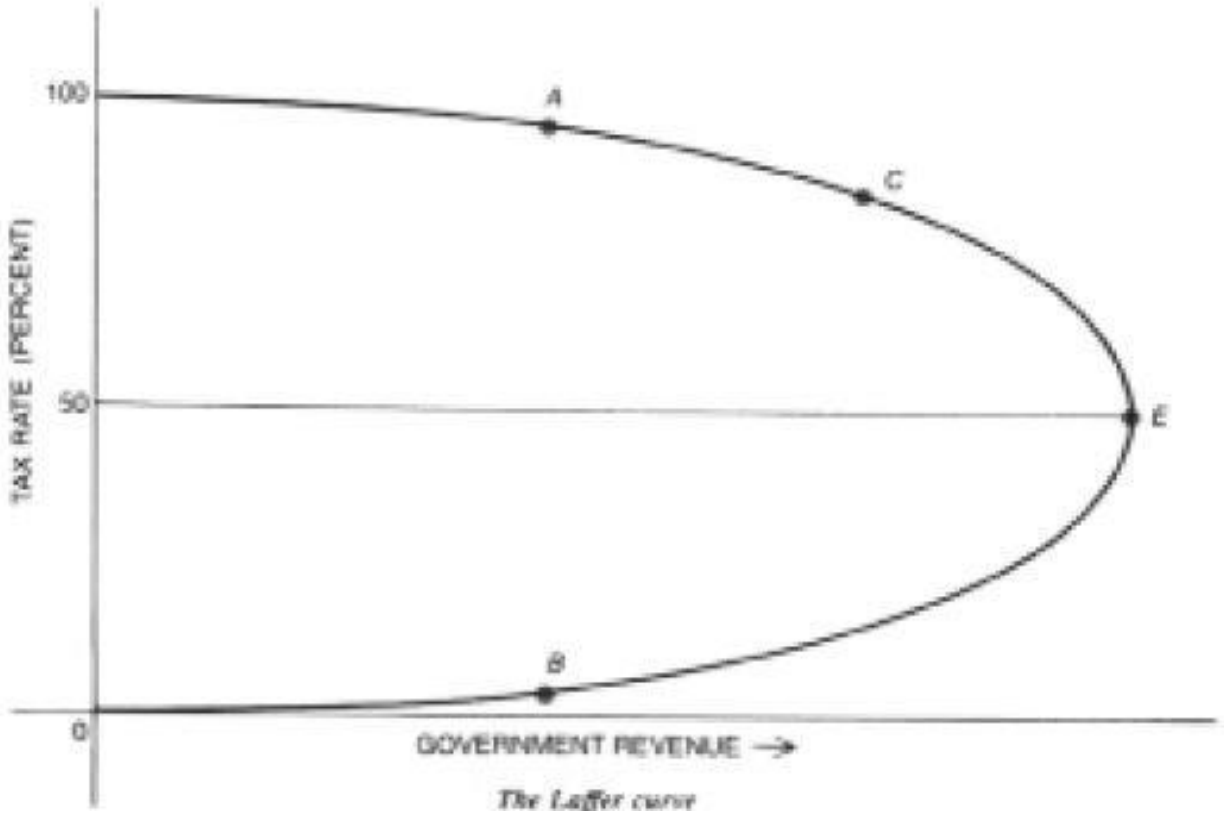
Czyż nie jest kwestią *geometrycznej konieczności*, że nasze świadome umysły znajdują się za *zakończeniem* wszelkich procesów *przychodzących*, a zarazem tuż przed *rozpoczęciem* wszelkich procesów *wychodzących*, realizujących nasze czynności? Przybywają z jednej granicy przez kanały wejściowe, na przykład z oka, potem idą nerwem optycznym przez różne ośrodki kory mózgowej i...? A podążając w drugą stronę, od drugiej granicy, przebiegając wzdłuż mięśni oraz neuronów motorycznych, które nimi sterują, docieramy do dodatkowej kory ruchowej i...? Te dwie trasy zbliżają się do siebie po dwóch zboczach, aferentnym (doprowadzającym) i eferentnym (odprowadzającym). Bez względu na to, jak trudne może być dokładne wyznaczenie granicy między tymi procesami w mózgu, to czy ze względów czysto geometrycznych nie musi istnieć najwyższy punkt, punkt zwrotny, punkt, w którym wszelkie operacje z jednej jego strony zachodzą *przed przeżyciem*, a wszelkie operacje z jego drugiej strony następują *po przeżyciu*?

Na rycinie Kartezjusza jasno widać, że wszystko dociera do szyszynki i z niej wychodzi. Mogłoby się zatem wydawać, że spoglądając na bardziej współczesny model mózgu, powinniśmy być w stanie oznaczyć kolorami te drogi, na przykład drogi aferentne na czerwono, a eferentne na zielono; tam, gdzie następowałaby zmiana koloru, tam znajdowałby się punkt graniczny między nimi.



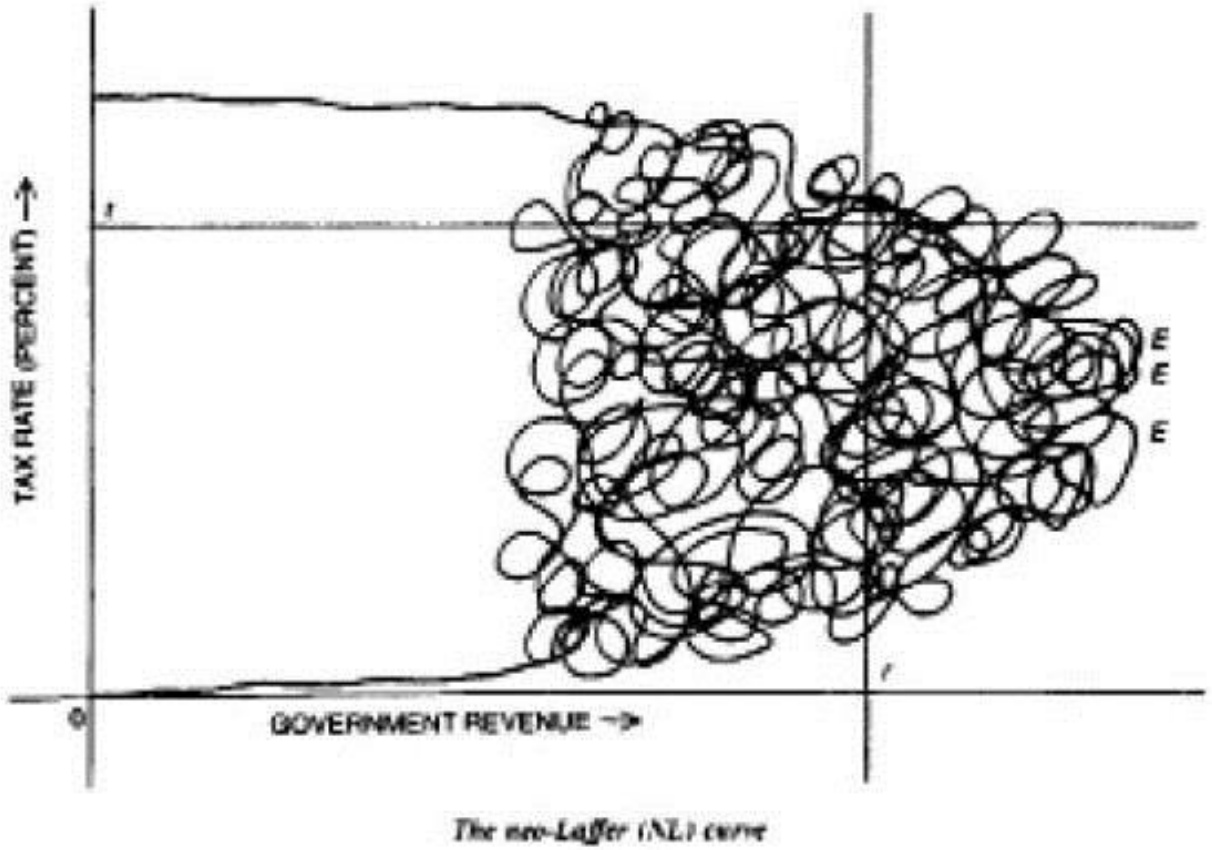
Ryc. 5.4

Ten osobliwie przekonujący argument może wam coś przypominać. Jest bratem bliźniakiem równie fałszywego argumentu, który ostatnio stał się zbyt znaczący: powszechnie znana krzywa Arthura Laffera, intelektualna podstawa (mówiąc swobodnie) reaganomiki[29]. Jeśli rząd nakłada 0 procent podatku, nie uzyskuje żadnych dochodów, a jeśli podatki wynoszą 100 procent, nikt nie dostanie pensji, więc nie będzie żadnych przychodów; jeśli są to 2 procenty, rząd uzyska mniej więcej dwa razy więcej dochodu niż przy 1 procencie i tak dalej, ale im wyższa jest stopa podatkowa, następować będą coraz mniejsze zyski; podatki staną się dokuczliwe. Lecz z drugiej strony 99 procent podatku to niewiele mniej niż 100 procent, więc nie nagromadzi się prawie żaden przychód; przy 90 procentach rząd zyska więcej, a jeszcze więcej przy 80 procentach. Stoki pokazanej krzywej mogą zniknąć, ale czy nie powinno istnieć, z geometrycznego punktu widzenia, miejsce, w którym krzywa zakręca, czyli wysokość podatku, która sprawia, że dochód będzie najwyższy? Idea Laffera była taka, że skoro obecna wysokość podatków wzrastała, obniżenie ich zwiększy przychód. Był to kuszący pomysł; wielu uważało, że tak właśnie musi być. Jednak, jak zauważył Martin Gardner, tylko dlatego, że końce krzywej są jasne, nie ma powodu, by nieznane części krzywej w regionach środkowych miały mieć gładki przebieg. Sarkastycznie proponuje on konkurencyjną „krzywą neo-Laffera", która ma więcej niż jedno „maksimum", a dostępność do każdego z nich jest uwarunkowana kompleksowością historii i warunków, której w żadnym razie nie może zdeterminować żadna zmiana pojedynczego czynnika (Gardner 1981).



The Laffer curve

Рис. 5.5



The neo-Laffer (NL) curve

Ryc. 5.6

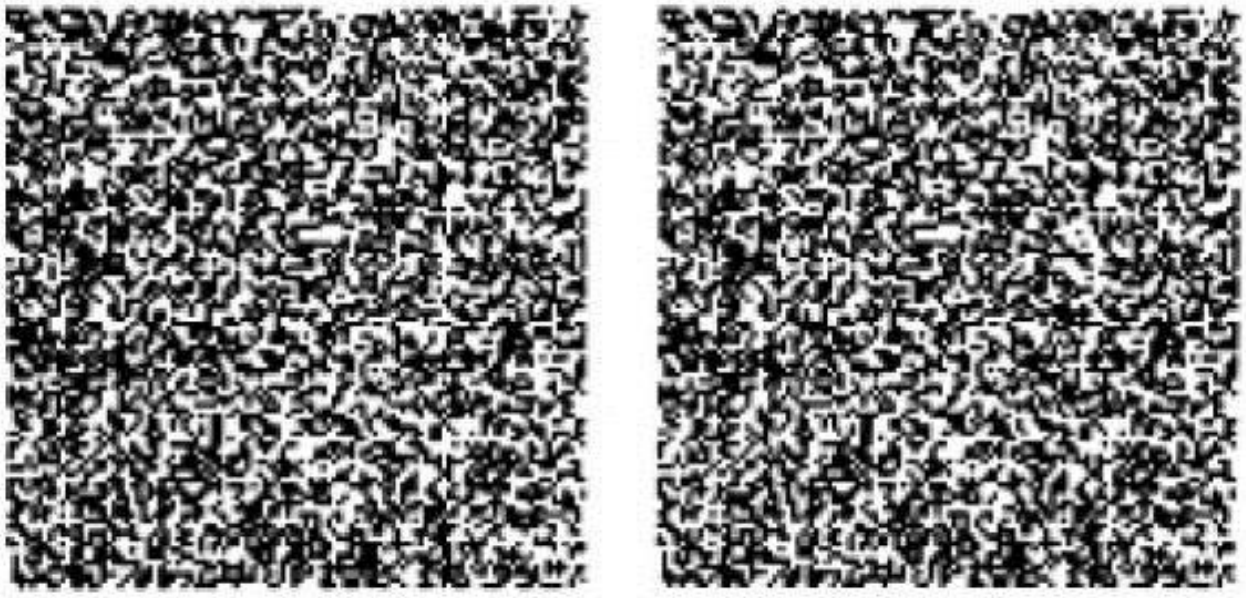
Powinniśmy wyciągnąć ten sam morał z tego, co leży we mgle między końcami dróg afarentnych i eferentnych: jasność krawędzi nie daje nam gwarancji, że to samo rozróżnienie będzie nadal obowiązywało *przez całą drogę*. „Techniczne poplątanie", które dla ekonomii przewiduje Gardner, jest i tak banalnie proste w porównaniu z mieszaniną czynności zachodzących w centralnych obszarach mózgu. Musimy przestać przypisywać mózgowi taki pojedynczy funkcyjny szczyt czy punkt centralny. Nie jest to niewinny skrót; to zły nawyk. Aby zerwać z takim nawykiem myślenia, musimy poznać pewne jego przypadki w akcji oraz znaleźć dobry pomysł, który może go zastąpić.

2. Pierwszy zarys modelu wielokrotnych szkiców

Oto pierwsza wersja takiego pomysłu, modelu wielokrotnych szkiców w świadomości. Spodziewam się, iż z początku będzie on wydawał się dość obcy i trudny do wyobrażenia – idea teatru kartezjańskiego jest aż tak głęboko w nas zakorzeniona. Według modelu wielokrotnych szkiców wszystkie rodzaje percepcji – oraz wszystkie rodzaje myśli czy aktywności umysłowej – są realizowane w mózgu przez równoległe, wielotorowe procesy interpretacji i przetwarzania bodźców zmysłowych. Informacje docierające do układu nerwowego są poddawane stałej „obróbce redakcyjnej". Na przykład ze względu na to, że głowa porusza się niewiele, a oczy – bardzo, obrazy na siatkówkach stale się zmieniają, niczym obrazy z domowych filmów kręconych przez osoby, które nie są w stanie utrzymać kamery w bezruchu. Jednak my tego tak

nie odbieramy. Ludzie często są zaskoczeni, gdy dowiadują się, że w normalnych warunkach ich oczy poruszają się szybkimi ruchami sakkadowymi, z prędkością około pięciu szybkich skoków na sekundę, oraz że ruch ten, tak jak ruch ich głów, jest wycięty we wczesnej fazie przetwarzania na drodze z gałki ocznej do... świadomości. Psychologowie dowiedzieli się wiele o mechanizmach pozwalających osiągać takie normalne efekty, ponadto odkrywając pewne efekty specjalne, takie jak interpretacja głębi w stereogramach składających się z przypadkowo ułożonych kropek (Julesz 1971). (Spójrz na rycinę 5.7).

Jeśli przez stereoskop[30] spojrzysz na te dwa odrobinę różniące się od siebie kwadraty (lub możesz po prostu popatrzeć na nie lekkim zezem, aby obrazki skleiły się w jeden – niektórzy potrafią to zrobić bez pomocy żadnego urządzenia), w końcu dostrzeżesz wyłaniający się, trójwymiarowy kształt, a stanie się to dzięki imponującemu procesowi redakcyjnemu w mózgu, który porównuje i zestawia ze sobą informacje z obu oczu. Odszukanie optymalnego zapisu może nastąpić bez uprzedniego rozpoznawania cech przedmiotów w zbiorze informacji. Wystarczająco dużo jest wyraźnych cech na niższym poziomie – pojedynczych punktów na stereogramie z przypadkowo ułożonymi kropkami – aby rozwiązanie się pojawiło.



Ryc. 5.7

Procesy redakcyjne mózgu potrzebują sporo czasu, aby osiągnąć tego rodzaju rezultaty, jednak inne efekty specjalne są natychmiastowe. Efekt McGurka (McGurk i Macdonald 1979) jest tego przykładem. Kiedy francuski film jest dubbingowany w języku angielskim, przez większość czasu widzowie nie są świadomi niezgodności między widzianymi ruchami warg a słyszanymi dźwiękami – chyba że dubbing jest zrobiony niechlujnie. Co się jednak dzieje, gdy stworzymy dźwięk odpowiadający obrazowi z wyjątkiem pewnych celowo pomylonych spółgłosek? (Wykorzystując naszego starego znajomego do nowych celów, możemy założyć, że usta filmowanej osoby mówią „z góry na dół", a dźwięk mówi „z dóry na gół"). Czego doświadczą odbiorcy? *Usłyszą* „z góry na dół". W sztucznie wywołanej rywalizacji redakcyjnej między oczami i uszami wygrywają oczy – w tym przypadku[31].

Owe procesy redakcyjne dokonują się w ułamkach sekundy, kiedy to mogą się pojawiać różnego rodzaju dodatki, uzupełnienia, wzbogacenia czy przejaskrawienia treści, w przeróżnej

kolejności. Nie przeżywamy bezpośrednio tego, co dzieje się na naszych siatkówkach, w naszych uszach, na powierzchni naszej skóry. Przeżywamy tylko wytwory wielu procesów interpretacji – będących w istocie procesami redakcyjnymi. Przejmują one stosunkowo nieprzetworzone, jednostronne reprezentacje i dostarczają reprezentacji porównanych, skorygowanych i ulepszonych. Procesy te przebiegają w strumieniach aktywności w różnych miejscach w mózgu. Ten fakt jest uznawany przez właściwie wszystkie teorie percepcji, jednak teraz jesteśmy już gotowi na nowatorską cechę modelu wielokrotnych szkiców: wykrycie czy rozpoznanie właściwości *musi nastąpić tylko raz*. To znaczy, że gdy dokonała się już „obserwacja" pewnej cechy w wyspecjalizowanym, zlokalizowanym obszarze mózgu, znajdujące się tam informacje nie muszą być wysyłane nigdzie indziej, aby zostały ponownie wyróżnione przez jakiegoś „głównego" wykrywacza. Innymi słowy, wykrycie nie prowadzi do *reprezentacji* tej właściwości przed publicznością teatru kartezjańskiego – gdyż teatr kartezjański nie istnieje.

Te przestrzennie i czasowo rozproszone w mózgu procesy ustalania treści można dokładnie zlokalizować, jednak ich początki nie są początkiem świadomości tych treści. Pozostaje pytaniem otwartym, czy pewna konkretna treść wyznaczona w taki sposób okaże się w końcu elementem świadomego przeżycia, a jak zobaczymy, pytanie, *kiedy coś staje się świadome*, jest błędne. Te rozproszone identyfikacje treści prowadzą z czasem do czegoś *podobnego* do strumienia czy ciągu narracyjnego, który można pojmować jako podlegający ciągłej redakcji przez wiele procesów rozproszonych w różnych częściach mózgu, które ciągle trwają. Ten strumień treści jedynie przypomina narrację ze względu na swoją różnorodność; w każdym momencie istnieje wiele „szkiców" fragmentów narracji na różnym etapie redakcji, w różnych miejscach w mózgu.

Sondowanie tego strumienia w różnych miejscach i momentach prowadzi do różnorodnych efektów i wywołuje najrozmaitsze narracje u osoby badanej. Jeśli sondowanie opóźni się za bardzo (powiedzmy: nastąpi następnego dnia), rezultatem może się okazać brak jakiejkolwiek narracji – bądź też narracja, która została uporządkowana lub „racjonalnie zrekonstruowana" aż do momentu, w którym brak jej spójności. Jeśli badamy „zbyt wcześnie", możemy zebrać informacje dotyczące tego, jak wcześnie dochodzi do konkretnego ustalenia treści w mózgu, jednak kosztem zmiany kierunku wielotorowego strumienia. Najważniejsze jest to, że model wielokrotnych szkiców unika kuszącego, choć błędnego założenia, iż musi istnieć pojedyncza narracja (którą można by nazwać „ostatecznym" czy „opublikowanym" szkicem), która jest obowiązująca – czyli jest *faktycznym* strumieniem świadomości osoby badanej, niezależnie od tego, czy eksperymentator (a nawet osoba badana) może uzyskać do niej dostęp.

Teraz jeszcze model ten prawdopodobnie nie jest szczególnie zrozumiały jako model świadomości, którą znasz ze swojego osobistego, bliskiego doświadczenia. To dlatego, że nadal jest ci wygodnie uznawać swoją świadomość za coś dziejącego się w teatrze kartezjańskim. Pozbycie się tego naturalnego, wygodnego zwyczaju i przedstawienie modelu wielokrotnych szkiców jako wyrazistej i wiarygodnej konkurencji będzie wymagało trochę pracy, czasem nawet dosyć dziwnej. Będzie to z pewnością najtrudniejsza część tej książki, jest ona jednak niezbędna do zrozumienia całej teorii i nie możemy jej pominąć! Na szczęście nie ma w niej matematyki. Musisz po prostu myśleć uważnie i jasno, upewniając się, że stwarzasz w głowie poprawną wizję i nie dajesz się skusić błędnym obrazom. Wyobraźni pomoże, jeśli po drodze pojawi się wiele łatwych eksperymentów myślowych. Przygotuj się zatem na intensywne ćwiczenia. Na koniec dostrzeżesz nową wizję świadomości, która zakłada istotną reformę (ale nie radykalną rewolucję) myślenia o mózgu. (Podobny model świadomości znajdziesz u Williama Calvina [1989] – „przędzenie scenariuszy").

Nową teorię łatwiej zrozumieć, widząc, jak radzi sobie ze stosunkowo łatwym

zjawiskiem, które podważa starą teorię. Dowód A to odkrycie związane z pozornym ruchem, o którym z radością mogę powiedzieć, że zostało wywołane pytaniem filozofa. Filmy i telewizja opierają się na tworzeniu pozornego ruchu przez prezentowanie szybko następujących po sobie „nieruchomych" obrazków, a od zarania dziejów kinematografii psychologowie badali owo zjawisko, nazwane *phi* przez Maxa Wertheimera (1912), który jako pierwszy zbadał je w sposób systematyczny. W najprostszym przypadku, jeśli co najmniej dwa małe punkty oddzielone od siebie o najwyżej 4 stopnie kątowe pola widzenia zostaną na moment podświetlone jeden po drugim, widać będzie jeden poruszający się punkt. Analizowano wiele przypadków phi, ale jeden z najbardziej zaskakujących został zrelacjonowany przez psychologów Paula Kolersa i Michaela von Grünau (1976). Filozof Nelson Goodman zapytał Kolersa, czy zjawisko phi zachodziłoby, gdyby punkty miały różne kolory, a jeśli tak, to co działoby się z kolorem punktu, gdy się porusza? Czy zniknęłaby iluzja ruchu, zastąpiona dwoma osobno świecącymi punktami? Czy iluzorycznie „poruszający się" punkt stopniowo zmieniłby jeden kolor na drugi, podążając torem między konkretnymi kolorami (trójwymiarowa kulka, która ukazuje wszystkie odcienie)? (Możesz spróbować stworzyć własne przypuszczenia, zanim przejdziesz do dalszej lektury). Gdy Kolers i Grünau wykonali eksperymenty, odpowiedź była niespodziewana: dwa punkty w innych kolorach świeciły się po 150 ms każdy (z 50-milisekundową przerwą); pierwszy punkt dawał wrażenie początku ruchu, a następnie gwałtownie zmieniał kolor *pośrodku iluzorycznego toru*. Goodman zastanawiał się: „jak to możliwe, że [...] uzupełniamy pośrednie czasoprzestrzenne położenia plamki wzdłuż pewnej linii zaczynającej się tam, gdzie wystąpił pierwszy błysk, a kończącej się tam, gdzie wystąpi drugi, *zanim jeszcze ten drugi ma miejsce*? (Goodman 1978/1997, s. 89–90).

To samo pytanie można oczywiście zadać o jakiekolwiek zjawisko phi, jednak kolorowe phi Kolersa wyraźnie pokazuje problem. Załóżmy, że pierwszy punkt jest czerwony, a drugi, przesunięty, zielony. Jeśli nie istnieje „prekognicja" w mózgu (ekstrawagancka hipoteza, którą odsuniemy na bok), iluzoryczna treść – *czerwony punkt przechodzący w zielony w połowie drogi* – nie może powstać dopiero *po* identyfikacji drugiego, zielonego punktu przez mózg. Jeśli jednak drugi punkt już znajduje się w „świadomym przeżyciu", czy nie byłoby za późno na wstawienie iluzorycznej treści między świadome przeżycie czerwonego punktu a świadome przeżycie punktu zielonego? W jaki sposób mózg dokonuje tej sztuczki?

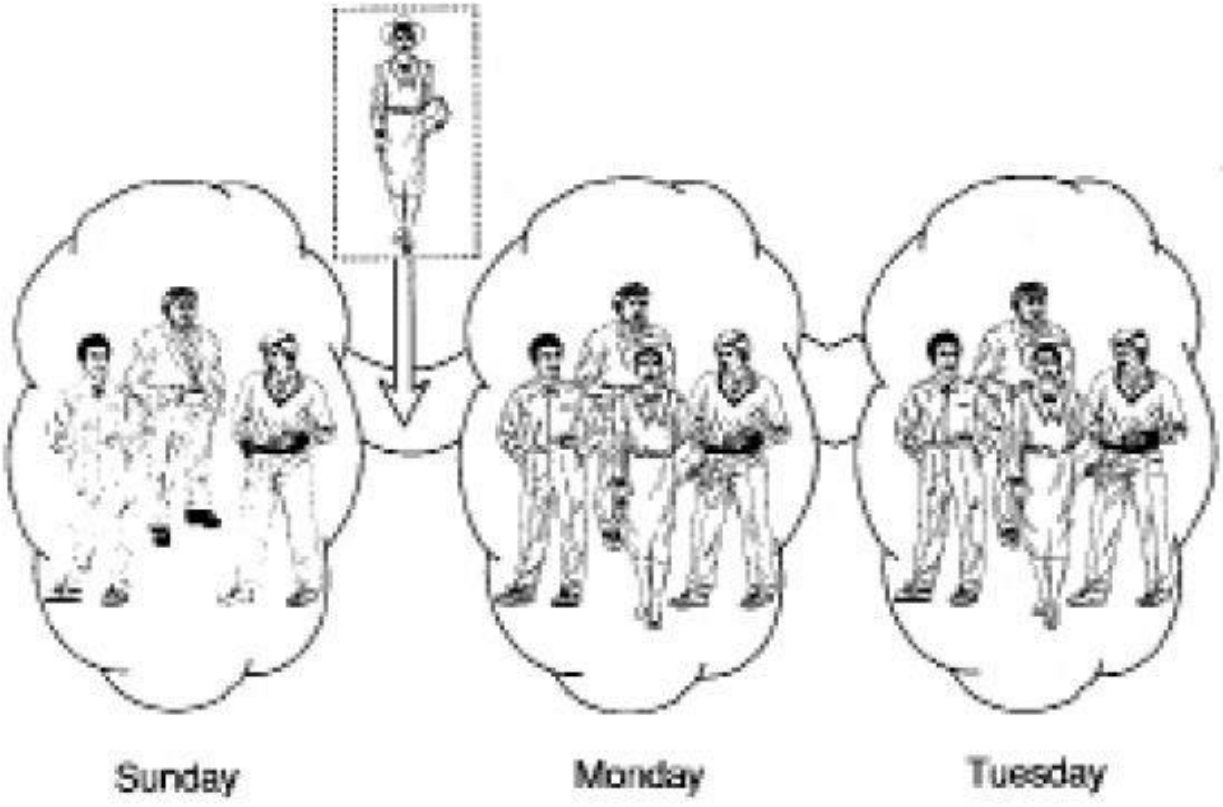
Zasada mówiąca, że przyczyny muszą poprzedzać skutki, ma zastosowanie do wielu rozproszonych procesów, realizujących redakcyjną pracę mózgu. Każdy proces, który potrzebuje informacji z jakiegoś źródła, musi w rzeczy samej na tę informację poczekać; dostanie się tam ona w swoim czasie. To właśnie wyklucza „magiczne" czy prekognitywne wyjaśnienia zmiany kolorów w zjawisku phi. Treść *zielony punkt* nie może zostać przydzielona do żadnego wydarzenia, świadomego bądź nie, dopóki światło z zielonego punktu nie dotrze do oka i nie spowoduje normalnej aktywności nerwowej w układzie wzrokowym do poziomu, w którym nastąpi rozróżnienie koloru zielonego. A zatem (iluzoryczne) rozróżnienie czerwonego przechodzącego w zielony musi nastąpić *po* rozróżnieniu zielonego punktu. Skoro to, co świadomie przeżywasz, to *najpierw czerwony, następnie czerwony przechodzący w zielony i w końcu zielony*, to „oczywiste" jest, że twoja świadomość całego zdarzenia musi być opóźniona do momentu po (nieświadomym?) dostrzeżeniu zielonego punktu. Jeśli to wyjaśnienie wydaje ci się kuszące, to cały czas jesteś w okowach teatru kartezjańskiego. Uciec z niego pomoże ci pewien eksperyment myślowy.

3. Modyfikacje orwellowskie i stalinowskie

Nie jestem pewien, czy inni nie potrafią mnie dostrzec, czy setną sekundy po tym, jak moja twarz pojawia się na ich horyzoncie, tysięczną sekundy po tym, jak spoglądają w moim kierunku, już zaczynają wymazywać mnie ze swojej pamięci: zapomniany, zanim stał się nieważny, smutny archanioł pamięci.

Ariel Dorfman, *Mascara*, 1988

Załóżmy, że ingeruję w twój mózg, wprowadzając do twojej pamięci fałszywą kobietę w kapeluszu w miejscu, gdzie nikt kapelusza na sobie nie miał (na przykład na niedzielnym przyjęciu). Jeśli w poniedziałek, gdy myślisz o przyjęciu, pamiętasz ją i nie znajdujesz żadnych wewnętrznych źródeł, które nakazywałyby ci wątpić w prawdziwość tego wspomnienia, nadal powiedzielibyśmy, że świadomego przeżycia tej kobiety nigdy *nie* było, a przynajmniej nie na niedzielnym przyjęciu. Oczywiście twoje późniejsze przeżycie (fałszywego) wspomnienia może być żywe i we wtorek możesz z pewnością zgodzić się, że masz wyraźne, świadome przeżycie dotyczące kobiety w kapeluszu, która była na przyjęciu, jednak będziemy się upierać, że *pierwsze* takie przeżycie pojawiło się w poniedziałek, nie w niedzielę (mimo że tobie wydaje się inaczej).



Ryc. 5.8

Nie mamy możliwości umieszczenia fałszywego wspomnienia metodami neurochirurgicznymi, jednak czasem nasze wspomnienia nas zawodzą, więc w tym przypadku to, czego nie możemy otrzymać chirurgicznie, samo następuje w mózgu. Niekiedy wydaje nam się, że pamiętamy coś, nierzadko wyraźnie, co nigdy się nie zdarzyło. Nazwijmy tego rodzaju poprzeżyciowe zanieczyszczenia czy modyfikacje pamięci „orwellowskimi", od przerażającej

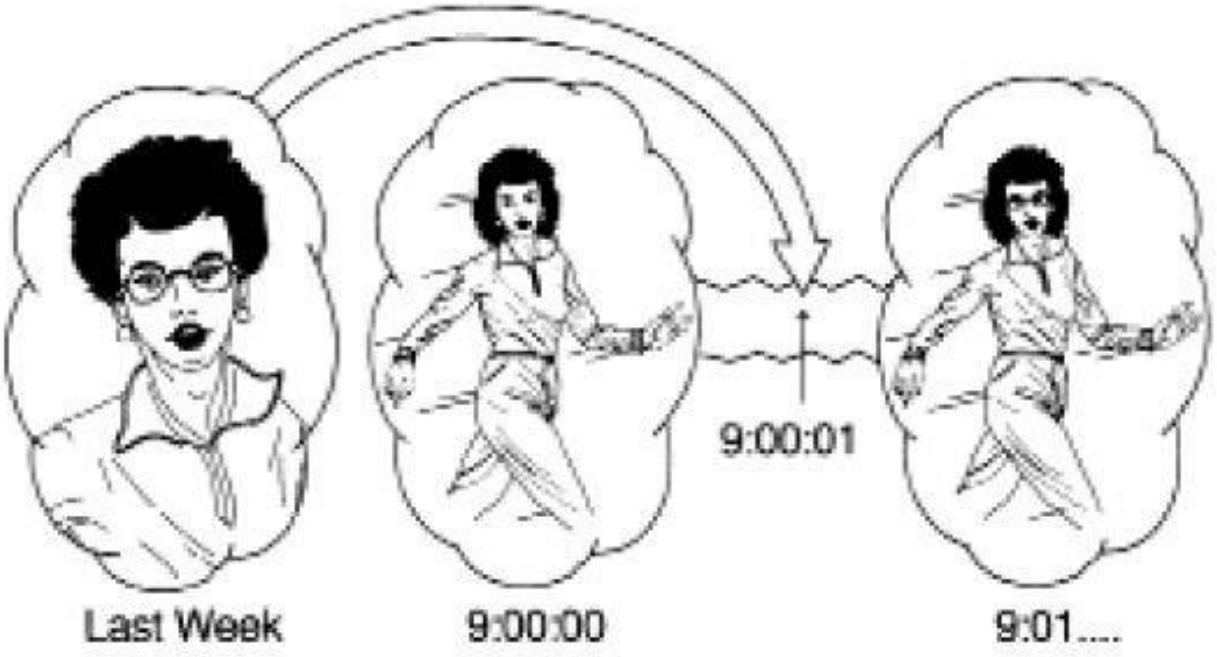
wizji Ministerstwa Prawdy z powieści George'a Orwella *1984*, pracowicie zmieniającego historię, uniemożliwiając późniejszy dostęp do (prawdziwej) przeszłości.

Możliwość tych poprzeżyciowych (orwellowskich) modyfikacji ujawnia aspekt jednego z naszych najbardziej fundamentalnych rozróżnień: rozróżnienie tego, co nam się wydaje i rzeczywistości. Ze względu na to, że dopuszczamy możliwość (przynajmniej teoretyczną) modyfikacji orwellowskiej, dopuszczamy też ryzyko przejścia od „oto co pamiętam" do „oto co się naprawdę wydarzyło", i dlatego czujemy opór – a mamy po temu słuszne powody – przed diabolicznym „operacjonizmem", który próbuje nas przekonać, że to, co pamiętamy (lub to, co znajduje się w historycznych archiwach), *po prostu jest* tym, co się faktycznie wydarzyło[32].

Modyfikacja orwellowska to jeden ze sposobów oszukania potomności. Inny to przygotowanie pokazowych procesów sądowych, opartych na szczegółowo napisanych scenariuszach fałszywych zeznań i przyznań się do winy, uzupełnionych o nieprawdziwe dowody. Tę taktykę nazwijmy „stalinowską". Zwróćmy uwagę na to, że gdy wiemy, który sposób falsyfikowania próbowano przeforsować, orwellowski lub stalinowski, jest to po prostu przypadek. Gdybyśmy przy okazji jakiejkolwiek *udanej* kampanii dezinformacyjnej mieli się zastanawiać, czy medialne relacje są orwellowskimi rekonstrukcjami zdarzeń, które nigdy nie zaistniały, czy też prawdziwymi relacjami z fałszywych procesów, które faktycznie się odbyły, moglibyśmy nie móc ich od siebie odróżnić. Gdyby *wszystkie* ślady – gazety, nagrania, osobiste wspomnienia, inskrypcje na nagrobkach, żyjący świadkowie – zniknęły z powierzchni ziemi lub zostały zmodyfikowane, w żaden sposób nie moglibyśmy się dowiedzieć, czy fałszerstwo nastąpiło *pierwsze*, a jego finałem było przedstawienie, którego szczegóły mamy przed sobą, czy może *po* jego wykonaniu fabrykacja historii zatarła czyn: żaden proces *faktycznie* się nie wydarzył.

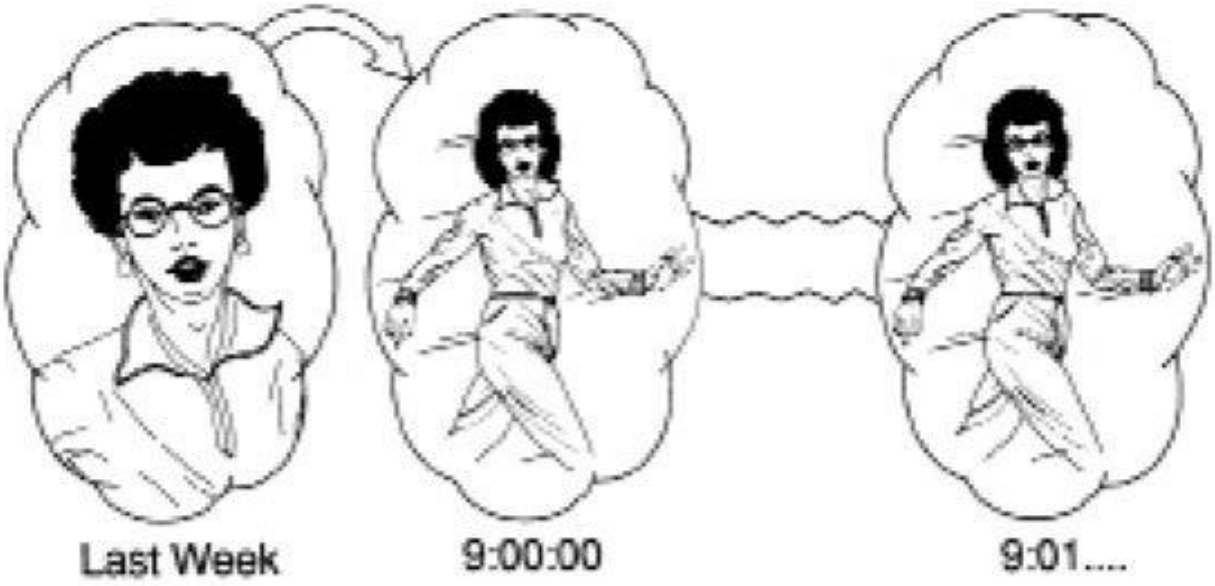
Różnica między orwellowską i stalinowską metodą wytwarzania mylących archiwów jest łatwo widoczna w makroskopowych skalach czasowych w codziennym życiu. Można by pomyśleć, że różnica ta jest widoczna *całkowicie*. Jednak jest to iluzja i możemy ją dostrzec w eksperymencie myślowym, który różni się od poprzedniego jedynie skalą czasową.

Wyobraź sobie, że stoisz na ulicy i przebiega koło ciebie długowłosa kobieta. Około sekundy *później* głębokie wspomnienie innej kobiety – krótkowłosej, w okularach – zanieczyszcza pamięć tego, co było przed chwilą przed twoimi oczyma: gdy chwilę później ktoś cię spyta o wygląd tej kobiety, mówisz, szczerze, choć błędnie, o jej okularach. Tak jak w przypadku kobiety w kapeluszu na przyjęciu, skłaniamy się ku temu, aby stwierdzić, że twoje pierwotne przeżycie *wzrokowe*, w przeciwieństwie do jego wspomnienia parę sekund później, nie było przeżyciem przedstawiającym kobietę w okularach. Jednak w rezultacie późniejszego zanieczyszczenia pamięci wydaje ci się bardzo wyraźnie, że w momencie gdy była widziana, miała okulary. Dokonała się modyfikacja orwellowska: był bardzo krótki moment, przed zanieczyszczeniem, w którym *nie* wydawało ci się, że ma ona okulary. W tym ulotnym momencie *rzeczywistością* twojego świadomego przeżycia była długowłosa kobieta *bez* okularów, ale ten historyczny fakt został anulowany; nie zostawił po sobie śladu dzięki zanieczyszczeniu pamięci, które nastąpiło w sekundę po zobaczeniu kobiety.



Ryc. 5.9

Takie rozumienie tego, co się wydarzyło, jest jednak ryzykowne, gdyż istnieje konkurencyjne wyjaśnienie. Twoje głęboko skryte, wcześniejsze wspomnienia kobiety w okularach mogły w równie prosty sposób zaburzać twoje przeżycie od samego początku, w trakcie przetwarzania informacji następującego „przed świadomością", w związku z czym mogły wystąpić *halucynacje* okularów od samego początku tego przeżycia. W takim wypadku twoje obsesyjne wspomnienie kobiety w okularach mogło cię oszukać w sposób stalinowski, tworząc przedstawienie w przeżyciu, które następnie poprawnie pamiętasz później dzięki zapisowi w twojej pamięci. Dla naiwnej intuicji te dwa przypadki są czymś zupełnie innym: w przypadku opowiedzianym na sposób pierwszy (Ryc. 5.9) nie było żadnych halucynacji w momencie, gdy kobieta przebiegła obok ciebie, ale halucynacje pojawiły się później; masz fałszywe wspomnienie faktycznego („prawdziwego") doświadczenia. Przypadek opowiedziany na drugi sposób (Ryc. 5.10) zakłada, że halucynacje się pojawiły, gdy kobieta przebiegła obok ciebie, a następnie prawidłowo pamiętasz te zwidy (które „naprawdę wydarzyły się w świadomości"). Czy na pewno są to dwie różne możliwości bez względu na to, na jak małe fragmenty podzielimy czas?



Ryc. 5.10

Nie. W tym przypadku rozróżnienie modyfikacji percepcyjnej i pamięciowej, które tak dobrze zdaje egzamin na większą skalę, nie gwarantuje nam sukcesu. Przesunęliśmy się w mgliste rejony, w których punkt widzenia osoby badanej jest przestrzennie i czasowo rozmyty, a pytanie „orwellowski czy stalinowski?" traci moc.

Istnieje okienko czasowe, które otwarło się w momencie, gdy długowłosa kobieta przebiegła obok ciebie, pobudzając twoje siatkówki, a zamknęło w momencie wyrażenia przez ciebie – przed samym sobą bądź przed kimś – końcowego przekonania o okularach. W którymś miejscu w tym przedziale czasu treść *okulary* została nieprawdziwie dodana do treści *długowłosa kobieta*. Możemy założyć (a moglibyśmy również kiedyś szczegółowo potwierdzić), że był krótki moment, w którym treść *długowłosa kobieta* została już rozróżniona przez mózg, jednak *zanim* treść *okulary* została błędnie do niego przyklejona. Z pewnością wiarygodne byłoby założenie, że owo rozróżnienie długowłosej kobiety przywołało zapis pamięciowy wcześniejszej kobiety w okularach. Nie wiedzielibyśmy jednak, czy owo nieprawdziwe doklejenie nastąpiło „przed czy po fakcie" – domniemanym fakcie „rzeczywistego, świadomego przeżycia". Czy najpierw pojawiła się świadomość długowłosej kobiety bez okularów, a następnie długowłosej kobiety w okularach, co byłoby późniejszym faktem, który wymazał z pamięci wcześniejsze przeżycie, czy może pierwszy moment świadomego przeżycia był od razu fałszywie powiązany z okularami?

Gdyby materializm kartezjański był prawdą, to pytanie musiałoby mieć odpowiedź, nawet jeśli my – i ty – nie moglibyśmy tego stwierdzić retrospektywnie żadnymi testami. A to dlatego, że treść, która „pierwsza dotarła do celu", to *długowłosa kobieta* lub *długowłosa kobieta w okularach*. Jednocześnie prawie wszyscy teoretycy twierdziliby, że kartezjański materializm jest błędny. Nie zauważyli jednak wynikającego stąd wniosku, że nie ma uprzywilejowanej linii mety, więc chronologiczna kolejność rozróżnień nie może ustalać subiektywnej kolejności przeżyć. Trudno przyjąć ten wniosek, ale może on stać się bardziej przekonujący, jeśli zbadamy trudności, które napotykamy, gdy upieramy się przy tradycyjnym podejściu.

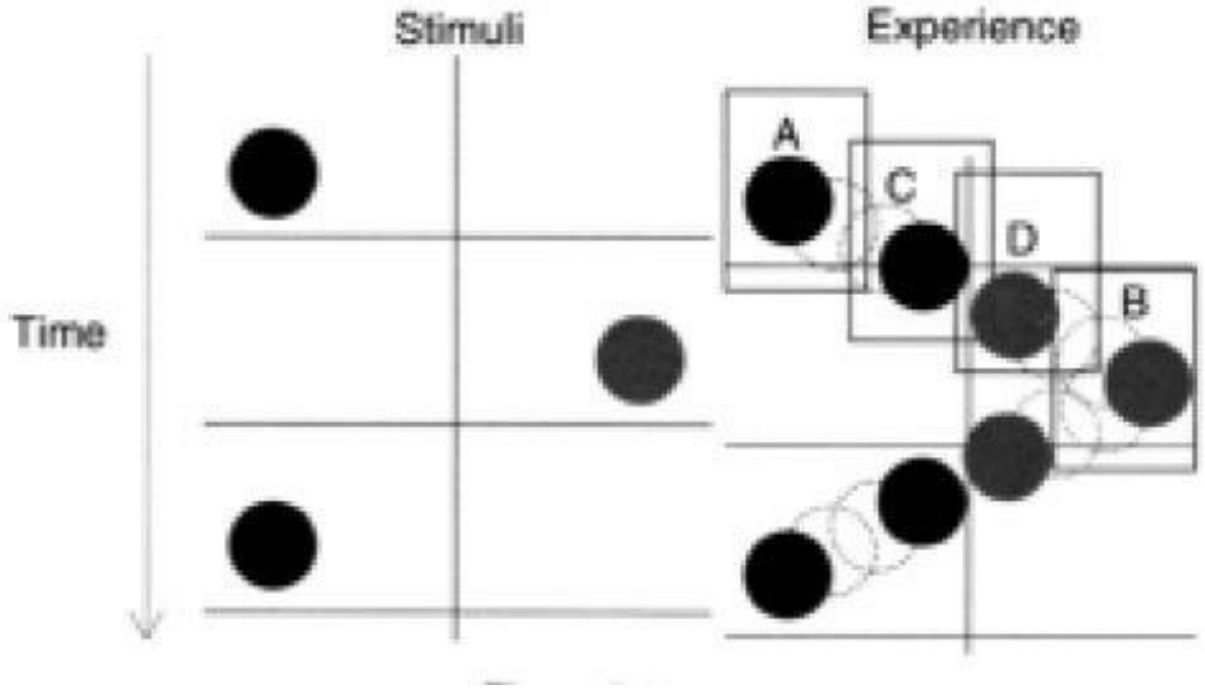
Spójrzmy na zjawisko phi Kolersa. Osoby badane *relacjonują* zmianę kolorów poruszających się punktów w połowie drogi między punktem czerwonym a zielonym. Ten

fragment tekstu został wyostrzony przez Kolersa za pomocą genialnie użytego wskaźnika, który osoby badane z opóźnieniem, ale najszybciej jak to możliwe „nakładały" na tor iluzorycznie poruszającego się punktu: umieszczając wskaźnik, dokonywały aktu mowy o treści: „Punkt zmienił kolor gdzieś *tutaj*" (Kolers i Grünau 1976, s. 330).

Zatem w heterofenomenologicznym świecie osób badanych zmiana koloru następuje w połowie drogi, a informacja o tym, jaki kolor się pojawia (i w jakim kierunku się porusza), musi się *skądś* brać. Przypomnij sobie słowa Goodmana o tej zagadce: „jak to możliwe, że [...] uzupełniamy pośrednie czasoprzestrzenne położenia plamki wzdłuż pewnej linii zaczynającej się tam, gdzie wystąpił pierwszy błysk, a kończącej się tam, gdzie wystąpi drugi, *zanim jeszcze ten drugi ma miejsce*?". Być może – jak sądzili niektórzy teoretycy – informacja pochodzi z *wcześniejszego przeżycia*. Być może, niczym pies Pawłowa spodziewający się jedzenia po każdym dzwonku, osoby badane zaczęły spodziewać się drugiego punktu po ujrzeniu punktu pierwszego i z przyzwyczajenia wykreślały tor, oczekując informacji o tym konkretnym przypadku. Jednak ta hipoteza została obalona. Nawet przy pierwszej próbie (czyli bez żadnych możliwości przewidywania) ludzie doświadczają zjawiska phi. Poza tym w kolejnych próbach kierunek oraz kolor drugiego punktu mogą być zmienione, a zjawisko nadal się pojawia. Zatem w jakiś sposób informacje o drugim punkcie (jego kolorze i lokalizacji) muszą być wykorzystane przez mózg do wytworzenia „zredagowanej" wersji, którą relacjonują osoby badane.

Zanalizujmy najpierw hipotezę, że istnieje mechanizm stalinowski: w redakcyjnym pokoju w mózgu, występującym przed świadomością, następuje zwłoka, opóźniająca pętla, przypominająca taśmowe opóźnienie używane w wyświetlanych na żywo programach, które daje cenzorom parę sekund na wycięcie nieprzyzwoitości, zanim sygnał zostanie wyemitowany. *Do pokoju redakcyjnego* najpierw przybywa klatka A z czerwonym punktem, a następnie, gdy przybywa klatka B z punktem zielonym, mogą powstać pewne klatki w lukach (C i D), które następnie będą wklejone do filmu (w kolejności A, C, D, B) na drodze do projekcji w teatrze świadomości. W momencie gdy „końcowy produkt" dociera do świadomości, ma już owe iluzoryczne uzupełnienia.

Istnieje konkurencyjna hipoteza mechanizmu orwellowskiego: krótko po uświadomieniu sobie pierwszego punktu *oraz* drugiego punktu (bez wrażenia ruchu) jakiś rewizjonistyczny historyk w stacji przechowującej mózgową bibliotekę pamięci zauważa, że w tym przypadku historia ta nie ma sensu, więc interpretuje te wydarzenia – czerwony, po którym następuje zielony – wymyślając narrację dotyczącą pośrednich momentów, zakładającą zmianę koloru w połowie drogi, i instaluje tę historię, dodając swój glosariusz, czyli klatki C i D (na Ryc. 5.11), w bibliotece pamięci do wglądu w przyszłości. Pracuje szybko, w ciągu ułamków sekundy – czas potrzebny na ustalenie (ale nie wypowiedzenie) werbalnej relacji z przeżycia – więc przechowywany w bibliotece pamięci zapis, na którym się opierasz, jest już zanieczyszczony. *Mówisz* i *wierzysz*, że widzisz iluzoryczny ruch i zmianę koloru, ale jest to tak naprawdę halucynacja pamięci, a nie autentyczny opis twojej pierwotnej świadomości.



5.11

Jak moglibyśmy sprawdzić, która z tych hipotez jest prawdziwa? Mogłoby się wydawać, że łatwo da się wykluczyć hipotezę stalinowską z powodu opóźnienia w świadomości, jakie zakłada. W eksperymencie Kolersa i von Grünaua była 200-milisekundowa różnica w zapaleniu czerwonego i zielonego punktu, a ponieważ, *na mocy założenia*, *całe* przeżycie nie może zostać wytworzone w pokoju redakcyjnym, dopóki treść *zielony punkt* nie dostanie się do niego, świadomość czerwonej kropki będzie musiała być opóźniona o co najmniej ten moment. (Gdyby pokój redakcyjny wysłał treść *czerwony punkt* do teatru świadomości bez zwłoki, zanim otrzymałby klatkę B i sfabrykowałby klatki C i D, osoba badana przypuszczalnie doświadczyłaby luki w filmie, opóźnienia o co najmniej 200 ms między A i C – zauważalnego na tyle, jak długa na sylabę przerwa jest w słowie, lub pięć brakujących klatek w filmie).

Załóżmy, że prosimy osoby badane o wciśnięcie przycisku, „gdy tylko doświadczą punktu". Odkrylibyśmy niewielką różnicę lub brak różnicy czasów reakcji na sam czerwony punkt i na ten sam punkt oraz doświadczony 200 ms później zielony punkt (w którym to przypadku osoby badane relacjonują zmianę koloru w pozornym ruchu). Czy dzieje się tak może dlatego, że w świadomości *zawsze* następuje przynajmniej 200-milisekundowe opóźnienie? Nie. Istnieje mnóstwo świadectw wskazujących, że reakcje pod świadomą kontrolą, jako że są wolniejsze niż reakcje takie jak odruchowe mrugnięcia, następują niemalże tak szybko, jak minimalne, fizycznie możliwe opóźnienia. Po odjęciu mierzalnego czasu drogi przychodzących i wychodzących ciągów impulsów oraz chwili na przygotowanie na reakcję nie ma czasu na „centralne redagowanie", w którym można by ukryć owe 200-milisekundowe opóźnienie. Zatem reakcja naciśnięcia przycisku musiała być rozpoczęta, zanim rozróżniono drugi bodziec, zielony punkt.

Wydaje nam się, że zwycięstwo powinno więc być przyznane hipotezie orwellowskiej, mechanizmowi modyfikacji poprzeżyciowej: w momencie gdy osoba badana uświadamia sobie czerwony punkt, rozpoczyna proces wciskania przycisku. *Podczas tego procesu* osoba ta staje się

świadoma zielonego punktu. *Następnie* oba te przeżycia zostają wymazane z pamięci i zastąpione w pamięci zrewidowanym nagraniem poruszającego się czerwonego punktu, który w połowie drogi zmienia kolor na zielony. Pewnie i szczerze, osoba badana *błędnie* relacjonuje dostrzeżenie czerwonego punktu przemieszczającego się w kierunku zielonego i zmieniającego kolor. Jeśli osoba badana usilnie twierdzi, że od samego początku była świadoma czerwonego punktu zmieniającego po drodze kolor, teoretyk orwellowski spokojnie wytłumaczy jej, że jest w błędzie; pamięć spłatała jej figla; fakt, że wcisnęła przycisk w tym momencie, jest ostatecznym dowodem na to, że była świadoma (nieruchomego) czerwonego punktu, zanim pojawił się punkt zielony. Dostała przecież polecenie wciśnięcia przycisku, *gdy uświadomiła sobie* czerwony punkt. Musiała być go świadoma około 200 ms, zanim mogła przeżyć jego ruch i zmianę koloru. Jeśli tak jej się wydaje, to po prostu się myli.

Obrońca alternatywy stalinowskiej nie jest jednak jeszcze pokonany. Tak naprawdę, twierdzi, osoba badana zareagowała na czerwoną kropkę, *zanim* stała się jej świadoma! Polecenie, które otrzymała (zareagować na czerwony punkt), jakoś przedostało się ze świadomości do pokoju redakcyjnego, co (nieświadomie) rozpoczęło proces wciskania przycisku, zanim zredagowana wersja (klatki ACDB) została wysłana do „wglądu". Pamięć osoby badanej nie spłatała jej figla; zdaje relację dokładnie z tego, czego była świadoma, z wyjątkiem świadomego wciśnięcia przycisku po zobaczeniu czerwonego punktu; jej „przedwczesne" wciśnięcie przycisku było wywołane nieświadomie (lub przedświadomie).

Tam, gdzie teoria stalinowska postuluje reakcję wciskania przycisku na nieświadome rozpoznanie czerwonego punktu, teoria orwellowska zakłada *świadome* przeżycie czerwonego punktu, które zostaje natychmiast zastąpione w pamięci przez swojego następcę. Oto nasza trudność: mamy dwa różne modele tego, co dzieje się w kolorowym zjawisku phi. Jeden zakłada stalinowskie „wypełnianie" na przedprzeżyciowej, oddolnej ścieżce, drugi zakłada orwellowską „modyfikację pamięci" na poprzeżyciowej, odgórnej ścieżce, a oba zgadzają się z tym, co mówi, myśli i pamięta osoba badana. Zwróćmy uwagę na to, że niemożność rozróżnienia tych dwóch wariantów nie odnosi się jedynie do obserwatorów *zewnętrznych*, o których można by przypuszczać, że brakuje im pewnych prywatnych danych, do których osoba badana ma „uprzywilejowany dostęp". Ty, jako osoba badana w eksperymencie ze zjawiskiem phi, *nie możesz* odkryć w doświadczeniu z twojej własnej perspektywy pierwszoosobowej niczego, co przemówiłoby za jedną czy drugą teorią; przeżycie byłoby takie samo w obu przypadkach.

Czy rzeczywiście tak jest? A gdyby zwrócić bardzo szczególną uwagę na swoje doświadczenie – czy na pewno nie można stwierdzić różnicy? Załóżmy, że eksperymentator chce ci to ułatwić i zwalnia pokaz, stopniowo wydłużając międzybodźcową lukę pomiędzy czerwonym i zielonym punktem. Oczywiste jest, że jeśli luka jest wystarczająco długa, zauważysz różnicę między faktycznym *odbiorem* ruchu i twoim *założeniem* ruchu. (Jest ciemna i burzowa noc; przy pierwszym uderzeniu błyskawicy widzisz mnie po swojej lewej stronie; dwie sekundy później błyska się znów i widzisz mnie po swojej prawej. Musiałem się poruszyć, wnioskujesz, i możesz z całą pewnością powiedzieć, że wyciągasz ten wniosek, ale nie było mnie widać, jak się poruszam). Gdy eksperymentator wydłuża przerwę pomiędzy bodźcami, nadejdzie moment, w którym rozpoczniesz dochodzić do tego rozróżnienia. Powiesz coś takiego:

Tym razem nie wydawało mi się, że czerwony punkt się ruszał, jednak po dostrzeżeniu zielonego punktu przyszło mi do głowy, że czerwony punkt się ruszył i zmienił kolor.

Tak naprawdę istnieje pośredni zakres przedziałów czasu, gdzie fenomenologia jest dosyć paradoksalna: widzisz punkty nieruchome i jednocześnie coś się porusza! Ten rodzaj pozornego ruchu łatwo odróżnić od szybszego, łagodnego rodzaju pozornego ruchu, który widzimy w filmach i telewizji, jednak nasza umiejętność przeprowadzania *tego* rozróżnienia jest nieistotna

w sporze pomiędzy teoretykami orwellowskimi i stalinowskimi. Zgadzają się oni, że możesz dokonać tego rozróżnienia w odpowiednich warunkach. Ale nie zgadzają się co do tego, jak opisać przypadki pozornego ruchu, którego *nie można* odróżnić od prawdziwego ruchu – od przypadków, w których rzeczywiście *dostrzegasz* pozorny ruch. Innymi słowy, czy w takich przypadkach twoja pamięć stroi sobie z ciebie żarty, czy to ty stroisz sobie żarty z siebie samego?

Jednak jeśli nawet ty, osoba badana, nie jesteś w stanie stwierdzić, czy to zjawisko stalinowskie, czy orwellowskie, to czy naukowcy – zewnętrzni obserwatorzy – nie mogliby znaleźć w twoim mózgu czegoś, co pokazałoby, które to zjawisko? Niektórzy wykluczają tę możliwość, twierdząc, że jest *niepojmowalne*. „Spróbuj sobie wyobrazić, że *ktoś inny* wie lepiej od ciebie, czego jesteś świadom! Niemożliwe!" Czy aby na pewno? Przyjrzyjmy się temu bliżej. Załóżmy, że owi naukowcy mają naprawdę precyzyjne dane (zgromadzone za pomocą różnorodnych technik skanowania mózgu) na temat dokładnego „czasu przybycia" czy „wytworzenia" każdej reprezentacji, każdego elementu treści, gdziekolwiek w twoim układzie nerwowym. To dałoby im *najwcześniejszy* moment, w którym osoba badana mogłaby w jakiś sposób zareagować – świadomie bądź nie – na pewną określoną treść (pomijając nadprzyrodzoną prekognicję). Jednak *rzeczywisty* moment, w którym uświadamiasz sobie tę treść (jeśli w ogóle), może nastąpić później. Trzeba by sobie ją uświadomić na tyle wcześnie, aby dało się wyjaśnić jej uwzględnienie w późniejszym akcie mowy – zakładając, że z definicji każdy element twojego heterofenomenologicznego świata jest elementem twojej świadomości. Wskaże to na *najpóźniejszy* moment, w którym treść „staje się świadoma". Jednak, jak widzieliśmy, jeśli pozostawia nam to jedynie kilkaset milisekund, podczas których musi nastąpić uświadomienie elementu, a jeżeli istnieje kilka różnych elementów, które muszą wystąpić w tym przedziale czasowym (czerwony i zielony punkt; długowłosa kobieta z okularami i bez), nie ma możliwości użycia twoich *raportów* do reprezentacji zdarzeń w świadomości.

Twoje późniejsze raporty werbalne muszą być neutralne w kwestii tych dwóch domniemanych możliwości, ale czy naukowcy nie mogą znaleźć jeszcze innych danych? Mogliby to zrobić, gdyby istniał dobry powód, aby twierdzić, że pewne zachowania niewerbalne (jawne bądź wewnętrzne) mogą być wyznacznikami świadomości. Takich powodów jednak nie ma. Obaj teoretycy zgadzają się, że nie istnieje reakcja behawioralna na treść, która to reakcja *nie mogłaby* być wyłącznie reakcją nieświadomą – z wyjątkiem późniejszej wypowiedzi. W modelu stalinowskim jest nieświadome wciskanie przycisku (a czemu nie?). Teoretycy zgadzają się również co do tego, że mogłoby istnieć świadome doświadczenie, które nie dawałoby żadnych efektów behawioralnych. W modelu orwellowskim istnieje chwilowa świadomość nieruchomego, czerwonego punktu, która nie pozostawia po sobie żadnego śladu w późniejszej reakcji (czemu nie?).

Oba modele mogą zgrabnie wyjaśnić *wszystkie* dane – nie tylko te, które już mamy, ale również te, których uzyskanie w przyszłości możemy sobie wyobrazić. Oba wyjaśniają raporty werbalne: jedna teoria mówi, że są niewinną pomyłką, a druga, że są rzeczywistymi raportami na temat przeżywanych błędów. Co więcej, możemy założyć, że jedni i drudzy mają *dokładnie* tę samą teorię na temat procesów zachodzących w twoim mózgu; zgadzają się co do miejsca i czasu, w których pomylona treść wkracza na ścieżki przyczynowe; nie zgadzają się jedynie co do tego, czy następuje to przed przeżyciem czy po nim. Dają to samo wyjaśnienie efektów niewerbalnych, z jednym małym wyjątkiem: jedna twierdzi, że są one wynikiem nieświadomie rozróżnionych treści, a druga, że są rezultatem świadomie rozróżnionych, ale zapomnianych treści. Wreszcie obie wyjaśniają subiektywne dane – to, co można zdobyć z perspektywy pierwszoosobowej – ponieważ zgadzają się nawet co do tego, jak osoby badane powinny to czuć: nie powinny być w stanie odróżnić błędnych przeżyć od tych natychmiast zapomnianych.

Zatem wbrew początkowym przypuszczeniom między tymi dwiema teoriami istnieje jedynie różnica werbalna (podobną diagnozę dają Reingold i Merikle 1990). Obie teorie przedstawiają dokładnie tę samą historię z wyjątkiem miejsca, w którym lokują mityczną granicę, momentu (a stąd miejsca), którego *szczegółowe* umiejscowienie jest jednocześnie neutralne w stosunku do wszystkich innych elementów teorii. Jest to różnica, która nie robi różnicy.

Przyjrzyjmy się współczesnej analogii. W świecie wydawniczym istnieje tradycyjne i zwykle dosyć bezkompromisowe rozróżnienie między redagowaniem tekstu przed jego wydaniem oraz poprawianiem błędów po publikacji, czyli erratą. Jednak w dzisiejszym świecie akademickim komunikacja elektroniczna przyspieszyła pewne procesy. W czasach edytorów tekstu, komputerowego składu i poczty elektronicznej często może być tak, że dostępnych jest kilka różnych szkiców artykułu, a autor wprowadza poprawki w odpowiedzi na komentarze otrzymane pocztą. Ustalenie momentu publikacji, a w związku z tym uznanie jednej z wersji artykułu za tekst *kanoniczny* – do którego będzie można odnosić się w bibliografii – staje się kwestią dość arbitralną. Często większość tych czytelników, dla których lektura tekstu ma duże znaczenie, czyta jedynie którąś z wczesnych wersji; ta „opublikowana" zostaje zarchiwizowana i nieaktywna. Jeśli więc szukamy ważnych efektów, większość ważnych efektów pisania artykułu do czasopisma, jeśli nie wszystkie, rozkłada się w kilku szkicach, a nie już po jego publikacji. Kiedyś było inaczej; właściwie wszystkie ważne efekty artykułu następowały *po* jego pojawieniu się w czasopiśmie oraz *z powodu* tego pojawienia się. Gdy już widzimy, że potencjalne „pasowanie" na publikację przestaje być funkcjonalnie istotne, to jeśli w ogóle potrzebujemy jakiegoś rozróżnienia tych sytuacji, będziemy musieli arbitralnie zdecydować, co należy uznać na opublikowanie tekstu. Nie ma naturalnej granicy czy punktu zwrotnego na ścieżce od szkicu do archiwum.

Analogicznie – i jest to fundamentalne założenie modelu wielokrotnych szkiców – jeśli ktoś chce uznać jakiś moment przetwarzania informacji w mózgu za moment pojawienia się świadomości, uczyni to arbitralnie. Zawsze możemy „zakreślić granicę" w strumieniu przetwarzania w mózgu, jednak nie ma funkcjonalnych różnic, które wskazałyby, że wszystkie poprzedzające je etapy i ustalenia były nieświadomymi lub przedświadomymi korektami, a wszelkie następujące potem zmiany w treści (ujawnione przez przypominanie) były poprzeżyciowymi zanieczyszczeniami pamięci. To rozróżnienie zanika, gdy zdarzenia następują tak blisko siebie.

4. Zrewidowany teatr świadomości

Złota zasada astronoma:
to, czego nie zapisano, nie istnieje.
Clifford Stoll, *Kukułcze jajo*, 1989/1998 [przeł. Tomasz Hornowski]

Jak powiada każda książka o prezentowaniu sztuczek magicznych, najlepsze triki są gotowe, zanim publika pomyśli, że się zaczęły. Być może teraz myślisz, że zrobiłem jakąś sztuczkę. Powiedziałem, że ze względu na czasoprzestrzenne rozmazanie punktu widzenia obserwatora w mózgu, żadne istniejące świadectwa nie pozwalają na rozróżnienie teorii świadomego przeżycia, orwellowskej i stalinowskiej, więc *nie ma między nimi żadnej różnicy*. Jest to pewien rodzaj operacjonizmu czy weryfikacjonizmu, który nie uwzględnia możliwości, że istnieją brutalne fakty, które są niedostępne dla nauki, nawet jeśli tą nauką jest heterofenomenologia. Poza tym wydaje się raczej oczywiste, że *istnieją* takie brutalne fakty – i że nasze bezpośrednie świadome przeżycia składają się z takich faktów!

Zgadzam się, że wydaje się to oczywiste; gdyby tak nie było, nie musiałbym tak się wysilać w tym rozdziale, by pokazać, że to, co jest tak oczywiste, jest tak naprawdę błędne. Zdaje się, że z premedytacją pominąłem coś analogicznego do wyśmianego kartezjańskiego teatru świadomości. Być może przypuszczasz, że pod przykrywką anty*dualizmu* („Pozbądźmy się tego strasznego świństwa!") przemyciłem coś, co do czego Kartezjusz miał rację: istnieje pewnego rodzaju funkcyjne miejsce, w którym elementy fenomenologiczne są... *wyświetlane*.

Czas, aby skonfrontować się z tym podejrzeniem. Nelson Goodman podejmuje ten temat, gdy mówi, że eksperyment kolorowego phi przeprowadzony przez Paula Kolersa zakłada, „że można wybierać jedynie między hipotezą uzupełniania retrospektywnego a hipotezą jasnowidztwa" (Goodman 1978/1997, s. 100). Musimy zrezygnować z jasnowidztwa, czym zatem jest „uzupełnianie retrospektywne"?

Nie przesądzając, czy percepcja pierwszego bodźca jest opóźniana, czy podtrzymywana, czy zapamiętywana, hipotezę tę nazywam doktryną wytwórczości retrospektywnej – tj. doktryną głoszącą, że to postrzegane jako zachodzące między bodźcami uzupełnianie dokonuje się nie wcześniej, nim zjawia się drugi bodziec. (Goodman 1978/1997, s. 98)

Z początku wydaje się, że Goodman waha się między teorią stalinowską (percepcja pierwszego punktu jest opóźniona) a orwellowską (percepcja pierwszego punktu zostaje podtrzymana czy zapamiętana), jednak ważniejsze jest to, że postulowana przez niego rewizjonistyczna hipoteza (orwellowski czy stalinowski) nie tylko dopasowuje osądy [badanych]; *wytwarza* materiał, aby *uzupełnić* luki:

wypełnia każdy punkt przejścia [...] wyłącznie jedną z dwu eksponowanych barw. (Goodman 1978/1997, s. 102)

Goodman nie dostrzega możliwości, że mózg tak naprawdę nie musi „wypełniać" czegokolwiek „wytworami" – ponieważ nikt nie patrzy. Jak wyjaśnia model wielokrotnych szkiców, gdy rozróżnienie zostało poczynione raz, nie musi być powtarzane; mózg po prostu dostosowuje się do wniosku, jaki wcześniej wyciągnął, inaczej interpretując informacje dostępne do regulacji późniejszego zachowania.

Goodman analizuje teorię, którą przypisuje Van der Waalsowi i Roelofsowi (1930), że „wrażenie ruchu tworzymy dopiero po spostrzeżeniu drugiego błysku i *rzutujemy* w przeszłość" (s. 90, podkr. D.C.D.). Sugeruje to pogląd stalinowski ze złowrogim dodatkiem: stworzona zostaje ostateczna wersja filmu, która zostaje wyświetlona na magicznym projektorze, a jego światło w jakiś sposób podróżuje w przeszłość na ekran umysłu. Bez względu na to, czy Van der Waals i Roelofs mieli na myśli właśnie to, gdy zaproponowali „wytwórczość retrospektywną", prawdopodobnie doprowadziło to Kolersa (1972, s. 184) do odrzucenia ich hipotezy i utrzymywania, że wszelka konstrukcja odbywa się w „czasie rzeczywistym". Dlaczego zatem w ogóle mózg miałby „wytwarzać" jakiekolwiek „wrażenie ruchu"? Dlaczego mózg nie może po prostu *stwierdzić, że nastąpił ruch* między punktami i wpleść tego retrospektywnego wniosku do strumienia przetwarzania? Czy to nie wystarczy?

Stop! To właśnie musi być sztuczka (jeśli chcemy to tak nazwać). Z perspektywy trzecioosobowej założyłem istnienie osoby badanej, badanej heterofenomenologicznie, pewnego rodzaj fikcyjnego „adresata", któremu my, zewnętrzni obserwatorzy, rzeczywiście poprawnie przypisywalibyśmy przekonanie, że ruch między punktami był przeżywany. Tak by się to *tej* osobie wydawało (a jest ona jedynie fikcją teoretyczną). Czy nie ma jednak *rzeczywistej* osoby, której mózg musi urządzać to całe przedstawienie, wypełniając wszystkie luki? To zapewne proponuje Goodman, gdy mówi o mózgu wypełniającym każdy punkt ścieżki. Ku czyjemu pożytkowi odbywa się ta cała animacja? Dla publiczności w teatrze kartezjańskim. Jednak *taka publiczność nie istnieje, gdyż nie istnieje teatr*.

Model wielokrotnych szkiców zgadza się z Goodmanem, że mózg retrospektywnie tworzy treści (osądy) o ruchu między punktami, które później są dostępne, regulując aktywność i odciskając swój ślad w pamięci. Jednak model wielokrotnych szkiców idzie dalej i twierdzi, że mózg nie zawraca sobie głowy „wytwarzaniem" jakichkolwiek reprezentacji, które miałyby „wypełniać" luki. Byłaby to strata czasu i (powiedzmy...) *farby*. Osąd *już* nastąpił, więc mózg może się zająć innymi sprawami[33].

„Rzutowanie w przeszłość" Goodmana jest wyrażeniem błędnym. Mogłoby znaczyć coś skromnego i do obronienia: mianowicie, że *odniesienie do pewnego momentu w przeszłości* jest zawarte w treści. W przypadku książki mogłoby to być zdanie: „powieść zabiera nas w podróż do starożytnego Rzymu...", którego nikt nie zinterpretowałby w metafizycznie ekstrawagancki sposób, twierdząc, że powieść to rodzaj maszyny do podróży w czasie. To skromne znaczenie jest spójne z innymi poglądami Goodmana, jednak Kolers najwyraźniej odebrał je jako coś metafizycznie radykalnego: że istnieje jakieś rzeczywiste rzutowanie pewnej rzeczy z jednego momentu w inny.

Jak zobaczymy w kolejnym rozdziale, zamieszanie wywołane tak radykalną interpretacją „rzutowania" negatywnie wpłynęło na interpretacje innych zjawisk. Ta sama dziwna metafizyka ciążyła na myśleniu o reprezentacjach przestrzeni. W czasach Kartezjusza Thomas Hobbes najwyraźniej sądził, że po dotarciu światła do oka wytwarza ono pewien ruch w mózgu, coś jest w jakiś sposób *odbijane* z powrotem do świata zewnętrznego.

> Przyczyną wrażenia zmysłowego jest ciało zewnętrzne, czyli przedmiot, który wywiera nacisk na organ właściwy dla danego zmysłu, bądź bezpośrednio, jak w smaku i dotyku, bądź pośrednio, jak w widzeniu, słyszeniu i wąchaniu. Ten nacisk za pośrednictwem nerwu i innych nici czy membran ciała przenosi się do wewnątrz w kierunku mózgu i serca, wywołuje tam pewien opór, nacisk przeciwny, czy też usiłowanie serca, by się od tego nacisku uwolnić. Usiłowanie to, ponieważ jest *skierowane na zewnątrz*, wydaje się być czymś leżącym na zewnątrz. (Hobbes 1651/1954, s. 9)

Uważał, że to przecież tam widzimy kolory – na powierzchni przedmiotów![34] W tym samym duchu ktoś mógłby założyć, że gdy wbijasz sobie nóż w palec u nogi, wywołuje to sygnały biegnące do „centrów bólu" w mózgu, które następnie wytwarzają „rzutowanie" bólu *z powrotem w dół do palca, czyli tam, gdzie powinien występować*. To przecież tam czujemy ból.

W latach pięćdziesiątych XX wieku ten pomysł był brany na tyle poważnie, że John Raymond Smythies, brytyjski psycholog, napisał artykuł, w którym szczegółowo go obalił[35]. Rzutowanie, o którym mówimy w kontekście takiego zjawiska, nie zakłada wywierania żadnego wpływu na zewnętrzną przestrzeń fizyczną i przypuszczam, że nikt już tak nie myśli. Jednak neurofizjologowie i psychologowie, a nawet akustycy projektujący stereofoniczne systemy głośników często mówią o tego rodzaju rzutowaniu i możemy zapytać, co tak naprawdę przez to rozumieją, jeśli nie coś, co zakłada fizyczną transmisję z jednego miejsca (czy momentu) w inny. Co się z tym wiąże? Przyjrzyjmy się prostemu przypadkowi:

> Dzięki rozmieszczeniu głośników stereo oraz balansie głośności każdego z nich słuchacz dokonuje *rzutowania* płynącego z nich dźwięku sopranu w punkt w połowie drogi między dwoma głośnikami.

Cóż to znaczy? Musimy do tego podejść uważnie. Jeśli głośniki wyją w pustym pokoju, rzutowanie nie zachodzi. Jeśli obecny jest słuchacz (obserwator ze zdrowymi uszami i mózgiem), „rzutowanie" się pojawia, jednak nie oznacza to, że słuchacz coś emituje w kierunku punktu w połowie drogi między głośnikami. Żadna właściwość fizyczna tego punktu i jego sąsiedztwa nie zostaje zmieniona przez obecność słuchacza. Krótko mówiąc, właśnie to mamy na myśli, gdy mówimy, że Smythies miał rację; nie istnieje projekcja żadnych cech wizualnych czy słuchowych

w przestrzeń. Co w takim razie się dzieje? Cóż, słuchaczowi *wydaje się*, że odgłos sopranu dochodzi z tamtego punktu. Co to *wydawanie się* zakłada? Jeśli odpowiemy, że zakłada „rzutowanie dźwięku przez obserwatora do tego miejsca w przestrzeni", wracamy oczywiście do punktu wyjścia, więc kusi nas, aby wprowadzić coś nowego, mówiąc na przykład tak: „obserwator dokonuje rzutowania dźwięku w *przestrzeni fenomenalnej*". Wygląda to na postęp. Zaprzeczamy, że projekcja zachodzi w przestrzeni fizycznej, zmieniliśmy jej lokalizację na przestrzeń fenomenalną.

Czym zatem jest przestrzeń fenomenalna? Czy jest to fizyczna przestrzeń w mózgu? Czy jest to przestrzeń na scenie teatru świadomości znajdującego się w mózgu? Niedosłownie. Metaforycznie? W poprzednim rozdziale widzieliśmy sposób rozumienia takich przestrzeni metaforycznych na przykładzie „obrazów umysłowych", którymi operował Shakey. W ścisłym, choć metaforycznym sensie Shakey rysował formy w przestrzeni, zwracał uwagę na konkretne miejsca w przestrzeni, wyciągał wnioski na podstawie tego, co dostrzegł w tych miejscach. Jednak ta przestrzeń była tylko przestrzenią *logiczną*. Była jak przestrzeń Londynu Sherlocka Holmesa, przestrzenią świata fikcyjnego, ale takiego, który systematycznie zakotwiczał się w rzeczywistych zdarzeniach fizycznych zachodzących w zwykłej przestrzeni w „mózgu" Shakeya. Jeśli potraktowalibyśmy wypowiedzi Shakeya jako jego „przekonania", moglibyśmy wówczas powiedzieć, że była to przestrzeń, na temat której Shakey *miał przekonania*, jednak nie sprawiłoby to, że stałaby się ona czymś rzeczywistym, tak jak czyjaś wiara w Feenomana nie sprawia, że Feenoman stał się prawdziwy. Oba przypadki są jedynie przedmiotami intencjonalnymi[36].

Istnieje zatem sposób rozumienia idei przestrzeni fenomenalnej jako przestrzeni logicznej. Jest to przestrzeń, do której czy w której nic nie dokonuje dosłownego rzutowania; jej właściwości są po prostu ustalane przez przekonania (heterofenomenologicznego) podmiotu. Gdy mówimy, że słuchacz dokonuje rzutowania dźwięku z jakiegoś miejsca w przestrzeni, chodzi nam *tylko* o to, że jemu wydaje się, że stamtąd właśnie dochodzi dźwięk. Czy to nie wystarczy? Czy może nie dostrzegamy „realistycznej" doktryny dotyczącej przestrzeni fenomenalnej, w której można dokonać rzutowania *rzeczywistego wydawania się*?

Dziś jesteśmy przyzwyczajeni do różnicy między lokalizacją przestrzenną nośnika przeżycia w mózgu a lokalizacją przeżywanego elementu „w przestrzeni przeżywanej". To znaczy, że rozróżniamy to, co reprezentuje, od tego, co jest reprezentowane, nośnik od treści. Zaszliśmy już tak daleko, że wiemy, iż wytwory percepcji wzrokowej nie są dosłownie obrazkami w głowie, mimo że *to, co reprezentują*, to coś, co obrazki pokazują bardzo wyraźnie: układ różnych widocznych cech w przestrzeni. Powinniśmy poczynić to samo rozróżnienie dla czasu: to, *kiedy* w mózgu następuje przeżycie, należy odróżnić od tego, kiedy wydaje się, że następuje. Rzeczywiście, jak zauważył psycholog Ray Jackendoff, tak naprawdę musimy zrozumieć tylko proste przedłużenie powszechnej wiedzy dotyczącej przeżycia przestrzeni. Reprezentacja przestrzeni w mózgu nie zawsze korzysta z przestrzeni w mózgu, a reprezentacja czasu w mózgu nie zawsze używa czasu w mózgu. Tak jak bezpodstawne jest istnienie przestrzennego rzutnika przezroczy, którego Smythies nie mógł znaleźć w mózgu, tak też nieuzasadniony jest czasowy rzutnik przezroczy, którego istnienie zakłada radykalne odczytanie „rzutowania w przeszłość" Goodmana.

Dlaczego ludzie czują potrzebę, aby zakładać istnienie rzutnika tego, co nam się wydaje? Dlaczego chętnie sądzą, że pokojom redakcyjnym w mózgu nie wystarczy jedynie wstawić treści do strumienia na drodze do zmiany zachowania i do pamięci? Być może dlatego, że chcą zachować odróżnienie rzeczywistości od pozoru w świadomości? Chcą się oprzeć diabolicznemu operacjonizmowi, który twierdzi, że to, co się zdarzyło (w świadomości), to tylko to, co

pamiętasz, że się zdarzyło. Dla modelu wielokrotnych szkiców „zapisanie" czegoś w pamięci jest kryterium dla świadomości; to jest to, *czym jest* dla „danej" bycie „odebraną" – odebraną w taki bądź inny sposób. Nie istnieje rzeczywistość świadomego przeżycia niezależna od efektów, które wywierają nośniki treści na późniejszą aktywność (i oczywiście na pamięć). Niebezpiecznie wygląda to na straszliwy operacjonizm, a być może kartezjański teatr świadomości jest po kryjomu czczony jako miejsce, gdzie cokolwiek, co dzieje się „w świadomości", dzieje się naprawdę, bez względu na to, czy jest to potem poprawnie zapamiętane. Załóżmy, że coś wydarzyło się w mojej obecności, ale pozostawiło we mnie swój ślad jedynie na „milionową część sekundy", jak w motcie autorstwa Ariela Dorfmana. Cóż może znaczyć powiedzenie, że byłem, nawet jeśli tylko chwilowo i nieskutecznie, tego świadom? Gdyby istniał gdzieś jakiś uprzywilejowany teatr kartezjański, mogłoby to przynajmniej znaczyć, że *film był tam fantastycznie przedstawiony*, nawet jeśli nikt nie pamięta, że go oglądał. (No właśnie!)

Teatr kartezjański może być pocieszającą wizją, gdyż zachowuje odróżnienie rzeczywistości od pozoru głęboko w ludzkim subiektywizmie, jednak oprócz tego, że jest ono naukowo bezzasadne, jest też metafizycznie wątpliwe, ponieważ tworzy dziwną kategorię czegoś obiektywnie subiektywnego – to, jakie rzeczywiście, obiektywnie wydają ci się rzeczy, nawet jeśli nie wydają ci się wydawać w ten sposób! (Smullyan 1981). Niektórzy myśliciele tak wojują z „weryfikacjonizmem" i „operacjonizmem", że przeczą mu nawet w jedynym miejscu, gdzie naprawdę ma sens: w sferze subiektywności. To, co Clifford Stoll nazywa „złotą zasadą astronoma", to sardoniczny komentarz do kaprysów pamięci oraz standardów dowodów naukowych. Staje się to jednak dosłowną prawdą, gdy myślimy o tym, co zostaje „zapisane" w pamięci. Moglibyśmy więc sklasyfikować model wielokrotnych szkiców jako *pierwszoosobowy operacjonizm*, gdyż obcesowo zaprzecza istnieniu w świadomości możliwości bodźca przy braku przekonania osoby badanej na temat tej świadomości[37].

Przeciwnicy tego operacjonizmu jak zwykle odwołują się do możliwych faktów poza zasięgiem testu operacjonalisty, jednak teraz operacjonalista jest właśnie podmiotem [świadomym], więc ta obiekcja odnosi odwrotny skutek: „Tylko dlatego, że nie potrafisz stwierdzić wybranymi przez siebie sposobami, czy masz świadomość *x-a*, czy nie, nie oznacza to, że tej świadomości nie było. Może świadomość *x-a wystąpiła*, ale po prostu nie możesz znaleźć na to dowodu!". Czy ktokolwiek, po chwili zastanowienia, naprawdę chce to powiedzieć? Domniemane fakty dotyczące świadomości znajdujące się poza zasięgiem „wewnętrznego" i „zewnętrznego" obserwatora to rzeczywiście fakty bardzo dziwne.

Trudno wykorzenić tę ideę. Spójrzmy, jak naturalnie brzmi: „Sądzę, że było tak, gdyż tak mi się wydawało". Łatwo odróżnić tutaj dwa różne stany czy zdarzenia: wydawanie się, że było właśnie tak, oraz późniejszy (i będący jego skutkiem) sąd, że było właśnie tak. Można pomyśleć, iż problem modelu wielokrotnych szkiców na przykład ze zjawiskiem phi jest taki, że nawet jeśli uwzględnia zjawisko polegające na sądzie osoby badanej dotyczącym ruchu między punktami, nie uwzględnia – a wyraźnie przeczy jego istnieniu – jakiegokolwiek zdarzenia mogącego zostać nazwanym pozornym ruchem między punktami, na którym ów sąd się „opiera". Gdzieś musi zostać „przedstawione świadectwo", nawet jeśli tylko w stalinowskim procesie pokazowym, aby ten sąd mógł być spowodowany przez to świadectwo lub przez nie uzasadniony.

Niektórzy przypuszczają, że taka intuicja ma uzasadnienie w fenomenologii. Wydaje im się, że rzeczywiście obserwują siebie osądzających w dany sposób *w rezultacie* tego, że owe rzeczy takie im się wydają. Nikt nigdy nie zaobserwował czegoś takiego „w ich fenomenologii", gdyż taki fakt dotyczący przyczyn byłby nie do zaobserwowania (jak dawno temu zauważył Hume)[38].

Zapytaj osobę badaną w eksperymencie ze zjawiskiem kolorowego phi: czy sądzisz, że

czerwony punkt poruszył się w prawo i zmienił kolor, ponieważ tak ci się wydawało, czy wydaje ci się, że się ruszył, ponieważ taki jest twój osąd? Załóżmy, że osoba badana daje rozwiniętą odpowiedź:

Wiem, że *w rzeczywistości* nie było poruszającego się punktu – to jedynie ruch pozorny – ale wiem też, że punkt *wydawał się* poruszać, więc poza moim osądem, że punkt zdawał się poruszać, istnieje zdarzenie, którego dotyczy mój osąd: wydający mi się ruch punktu. Nie było żadnego prawdziwego ruchu, więc musiało istnieć prawdziwe wydawanie się, że jest ruch, żeby mój osąd mógł się pojawić.

Być może teatr kartezjański jest popularny, gdyż jest miejscem, gdzie coś, co się nam wydaje, może nastąpić – osobno od oceniania. Jednak przedstawiony właśnie, rozwinięty argument jest błędny. Postulowanie „czegoś rzeczywistego" oprócz osądu lub „ujmowania" wyrażonego w raporcie osoby badanej mnoży byty ponad konieczność. Gorzej, mnoży byty ponad możliwość; rodzaj wewnętrznego przedstawienia, w którym następuje to, co naprawdę nam się wydaje, jest beznadziejnym metafizycznym unikiem, sposobem na to, aby mieć ciastko i zjeść ciastko, zwłaszcza dlatego, że ci, którzy mówią w ten sposób, chętnie upierają się przy tym, że to wewnętrzne przedstawienie nie odbywa się w jakiś tajemniczy, dualistyczny sposób, w którym przestrzeń przenika się z kartezjańskim eterem duchowym. Jeśli odrzuca się dualizm kartezjański, trzeba tak naprawdę odrzucić przedstawienie, które trwałoby w teatrze kartezjańskim, oraz publiczność, gdyż ani przedstawienia, ani publiczności nie można znaleźć w mózgu, a mózg jest jedynym realnym miejscem, w którym można ich szukać.

5. Model wielokrotnych szkiców w akcji

Przyjrzyjmy się ponownie modelowi wielokrotnych szkiców, nieco go rozbudowując i skupiając się trochę dokładniej na sytuacji w mózgu, która jest jego podstawą. Dla ułatwienia skoncentruję się na tym, co dzieje się w mózgu podczas przeżycia wzrokowego. Następnie rozbuduję model, aby uwzględnić w nim inne zjawiska.

Bodźce wzrokowe wywołują ciągi zdarzeń w korze, polegających na stopniowym rozróżnianiu coraz większych szczegółów. W różnych momentach i miejscach podejmowane są różne „decyzje" lub powstają „osądy"; z tego powodu obszary mózgu przechodzą w stany rozróżniania różnych własności, na przykład samo początkowe wystąpienie bodźca, następnie jego lokalizacja, kształt, następnie kolor (inną ścieżką przetwarzania), (pozorny) ruch, aż w końcu następuje rozpoznanie obiektu. Te stany rozróżniania, zlokalizowane w różnych miejscach, wywołują efekty jeszcze gdzie indziej, biorąc udział w kolejnych rozróżnieniach i tak dalej (Van Essen 1979; Allman, Meizin i McGuinness 1985; Livingstone i Hubel 1987; Zeki i Shipp 1988). Naturalne, choć naiwne, pytanie brzmi następująco: Gdzie to wszystko się ze sobą łączy? Odpowiedź brzmi: nigdzie. Niektóre rozproszone stany reprezentacyjne w różnych obszarach szybko zanikają, nie pozostawiając po sobie żadnych dalszych śladów. Inne pozostawiają ślady w raportach werbalnych z przeżyć oraz w pamięci, w „gotowości semantycznej" oraz innego rodzaju nastawieniach percepcyjnych, w stanach emocjonalnych, skłonnościach behawioralnych i innych. Niektóre takie efekty – na przykład wpływ na raporty werbalne – są przynajmniej symptomami świadomości. Nie ma jednak jednego miejsca w mózgu, przez które muszą przejść wszystkie te ciągi przyczynowe, aby pozostawić swoją treść „w świadomości".

W momencie gdy dojdzie do jakiegoś rozróżnienia, może być ono dostępne w ramach regulacji jakiegoś zachowania, na przykład wciśnięcia przycisku (lub uśmiechu czy komentarza), a także w ramach modulowania jakiegoś wewnętrznego stanu informacyjnego. Na przykład

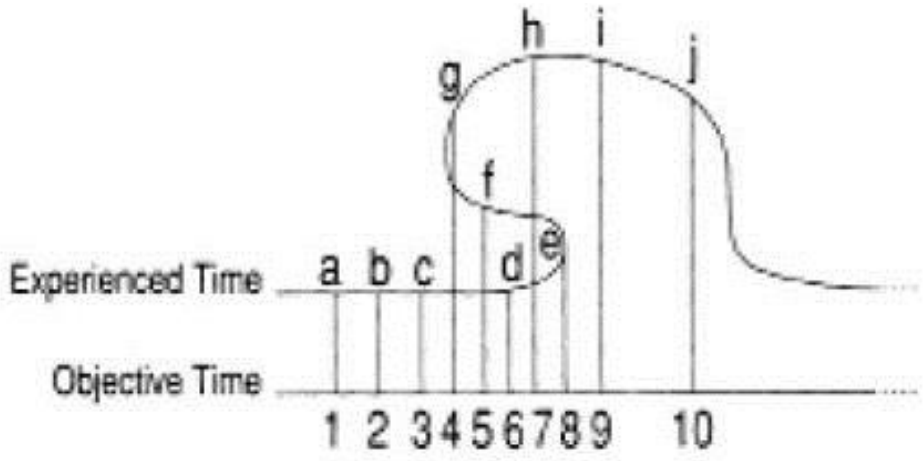
rozróżnienie zdjęcia kolczyka może stworzyć „nastawienie percepcyjne" – sprawiając, że chwilowo łatwiejsze będzie dostrzeganie uszu (a nawet części ciała) na innych zdjęciach – lub może aktywować konkretny obszar semantyczny, sprawiając, że chwilowo bardziej prawdopodobne będzie przeczytanie słowa „ucho" jako oznaczenia części ciała, a nie uchwytu dzbanka. Jak już zauważyliśmy, ten wielotorowy proces następuje w ciągu ułamków sekundy, w którym to czasie mogą pojawić się w różnej kolejności przeróżne dodatki, uzupełnienia, poprawki czy zmiany treści. Te natomiast po pewnym czasie dostarczają czegoś *na kształt* narracyjnego strumienia lub ciągu, a ten można traktować jako przedmiot ciągłej edycji, której jest on poddawany przez wiele procesów w różnych obszarach mózgu. Ten proces trwa bez końca i ciągnie się w przyszłości. Pojawiają się treści, które poddawane są korekcie, wpływając na interpretacje innych treści lub na zmiany w zachowaniu (werbalnym i innym), a jednocześnie pozostawiają w pamięci ślady, które w końcu zanikają bądź zostają częściowo lub całkowicie wchłonięte czy zastąpione przez późniejsze treści. Taka plątanina treści jest jedynie czymś na kształt narracji ze względu na swoją wielorakość; w jakimkolwiek momencie istnieją wielokrotne szkice fragmentów narracji, na różnych etapach przetwarzania w różnych miejscach w mózgu. Niektóre z tych treści w pewnym stopniu wpłyną na dalsze przetwarzanie, a następnie znikną bez długotrwałego efektu, a niektóre nie wpłyną na nic; inne zaś nadal będą odgrywały przeróżne role w dalszym przekształcaniu wewnętrznego stanu oraz zachowania. Garstka będzie na tyle trwała, że zaznaczy swoją obecność w komunikatach prasowych wydanych w formie zachowania werbalnego.

Analiza tego strumienia w różnych momentach prowadzi do różnorakich efektów, ujawniających różnorakie narracje – a są to właśnie narracje: pojedyncze wersje wycinków ze „strumienia świadomości". Jeśli badanie przeprowadzimy zbyt późno, w rezultacie możemy nie otrzymać żadnej narracji. Jeśli sondujemy zbyt wcześnie, możemy zebrać informacje o tym, jak wcześnie w strumieniu dokonuje się konkretne rozróżnienie, jednak jednocześnie przerywając zwykły rozwój strumienia.

Czy istnieje „optymalny moment badania"? Zakładając, dosyć wiarygodnie, że po jakimś czasie tego rodzaju narracje stopniowo giną zarówno z powodu zanikających detali, jak i ich narcystycznego upiększania (to, co powinienem był powiedzieć na przyjęciu, często zamienia się w to, co rzeczywiście powiedziałem), możemy przeprowadzić uzasadnioną analizę zaraz po interesującej nas sekwencji bodźców. Jednak chcemy też uniknąć ingerencji w zjawisko badaniem przedwczesnym. Percepcja niezauważalnie zamienia się w pamięć, a „natychmiastowa" interpretacja niepostrzeżenie zamienia się w racjonalną rekonstrukcję, zatem nie ma jednego jedynego punktu dla wszystkich kontekstów, na którym można by skupić badanie.

To, czego jesteśmy świadomi w danym okresie, nie jest zdefiniowane niezależnie od badań wyzwalających narracje o tym czasie. Ze względu na to, że narracje te są poddawane bezustannej rewizji, nie istnieje jedna, którą moglibyśmy uznać za wersję obowiązującą, za „pierwsze wydanie", gdzie znajdziemy wydarzenia, które nastąpiły w strumieniu świadomości osoby badanej – a wszelkie odchylenia od tej wersji muszą być zanieczyszczeniami tekstu. Jednak każda narracja (bądź jej fragment), która zostanie wywołana, daje „oś czasu", subiektywny ciąg zdarzeń z punktu widzenia obserwatora. Oś ta może być następnie porównana z innymi osiami, szczególnie z obiektywnym ciągiem zdarzeń w mózgu tego obserwatora. Jak widzieliśmy, te dwie linie *mogą* nie pokrywać się ściśle ze sobą: mimo że (błędne) rozróżnienie *czerwonego przechodzącego w zielony* nastąpiło w mózgu *po* rozróżnieniu zielonego punktu, *subiektywny* lub *narracyjny* ciąg to oczywiście *czerwony punkt, następnie czerwony zamieniający się w zielony, a w końcu zielony punkt*. Zatem w ramach wycinka czasu z punktu widzenia osoby

badanej mogą pojawić się różnice w kolejnościach, prowadzące do osobliwości.



Ryc. 5.12

Nie ma nic metafizycznie ekstrawaganckiego czy wymagającego w błędnej rejestracji [kolejności zdarzeń][39]. Jest ona nie mniej tajemnicza czy antyprzyczynowa niż uzmysłowienie sobie, że pojedyncze sceny w filmach często są kręcone w innej kolejności, lub tego, że gdy czytasz zdanie „Bill dotarł na przyjęcie szybciej niż Sally, ale Jane dotarła przed nimi obojgiem", dowiadujesz się o przybyciu Billa, zanim przeczytasz o wcześniejszym przybyciu Jane. Przestrzeń i czas reprezentującego to jeden układ odniesienia; przestrzeń i czas tego, co jest reprezentowane, to coś innego. Jednak ten metafizycznie nieszkodliwy fakt mimo wszystko ugruntowuje pewną fundamentalną kategorię metafizyczną: gdy fragment świata pojawia się i tworzy gmatwaninę narracji, tym fragmentem świata jest obserwator. To na tym polega bycie obserwatorem, to, jak to jest być czymś.

Jest to przybliżony szkic mojego konkurencyjnego modelu. To, jak bardzo różni się on od modelu teatru kartezjańskiego, muszę jeszcze dokładniej objaśnić, pokazując, jak wyjaśnia konkretne zjawiska. W następnym rozdziale za pomocą tego modelu rozpracujemy trudne zagadnienia, najpierw jednak pokrótce zastanówmy się nad pewnymi zwykłymi i znajomymi przykładami, często dyskutowanymi przez filozofów.

Prawdopodobnie wiesz z własnego doświadczenia, że można prowadzić samochód przez wiele kilometrów, jednocześnie rozmawiając (lub cicho monologując), a następnie zdać sobie sprawę, że absolutnie nie pamięta się drogi, ruchu ulicznego, wykonywania czynności związanych z prowadzeniem samochodu. To tak, jakby prowadził ktoś inny. Wielu teoretyków (przyznaję, że w tym i ja – Dennett 1969, s. 166 i nast.) uznawało to zjawisko za wspaniały przykład „nieświadomej percepcji i inteligentnego działania". Czy jednak *rzeczywiście* nie było w twojej świadomości tych wszystkich mijających cię samochodów, świateł, zakrętów na drodze? Twoja uwaga była zwrócona gdzie indziej, ale po spytaniu cię, co było widać w różnych momentach jazdy, z pewnością możesz być w stanie zrelacjonować przynajmniej jakieś niepełne szczegóły. Zjawisko „nieświadomej jazdy" lepiej wyjaśnia się jako przypadek falującej świadomości, przerywanej szybkimi zanikami pamięci.

Czy cały czas masz świadomość tykania zegara? Jeśli nagle przestanie chodzić, zauważysz to i natychmiast będziesz potrafił stwierdzić, co się zatrzymało; tykanie, „którego nie masz świadomości" aż do momentu, w którym ustało, i którego „nigdy nie byłbyś świadomy",

gdyby nie ustało, teraz wyraźnie jest w twojej świadomości. Jeszcze bardziej uderzającym przykładem jest zjawisko, w którym jesteś w stanie liczyć, retrospekcyjnie w pamięci przeżycia, uderzenia zegara, którego wybijanie zostało zauważone dopiero po czterech czy pięciu uderzeniach. W jaki sposób możesz tak jasno *pamiętać słyszenie* czegoś, co w ogóle nie było w twojej świadomości? To pytanie zdradza przywiązanie do modelu kartezjańskiego; nie ma sztywnych faktów dotyczących strumienia świadomości, niezależnych od konkretnych pytań i badań.

Rozdział 6

Czas i przeżycie

Mogę wprawdzie powiedzieć: moje przedstawienia następują po sobie, ale to znaczy tylko, że jesteśmy świadomi ich jako następujących w kolejności czasowej, tzn. wedle formy zmysłu wewnętrznego.

Immanuel Kant, *Krytyka czystego rozumu*, 1781
[przeł. Roman Ingarden]

W poprzednim rozdziale widzieliśmy w zarysie, jak model wielokrotnych szkiców likwiduje problem „rzutowania w przeszłość", pominęliśmy jednak pewne spore komplikacje. W tym rozdziale zajmiemy się tymi zagadnieniami na bardziej wymagającym terytorium, badając i rozwiązując kilka sporów powstałych wśród psychologów i neuronaukowców, dotyczących odpowiedniego wyjaśnienia pewnych wciąż niedających nam spokoju eksperymentów. Myślę, że możliwe jest zrozumienie reszty książki bez podążania za argumentami przedstawionymi w tym rozdziale, więc może on być pominięty lub jedynie przejrzany, lecz starałem się zaprezentować tę kwestię wystarczająco jasno dla laików i mogę wymienić sześć dobrych powodów, dla których warto się w ten rozdział zagłębić.

(1) Jest jeszcze wiele niejasnych kwestii w moim modelu wielokrotnych szkiców, a dzięki spojrzeniu, jak radzi on sobie z wyjaśnianiem pewnych zagadnień, łatwiej zrozumieć, jaka jest jego struktura.

(2) Jeśli masz jeszcze jakieś wątpliwości co do tego, jak model wielokrotnych szkiców różni się jako teoria empiryczna od teatru kartezjańskiego, zostaną one rozwiane za pomocą kilku widowiskowych starć.

(3) Jeśli myślisz, że krytykuję słomianą kukłę, otuchy doda ci odkrycie, iż pewni eksperci są wielce zaniepokojeni, gdyż wbrew sobie są prawdziwymi kartezjańskimi materialistami.

(4) Jeśli podejrzewasz, że mój model oparłem na jednym, specyficznym zjawisku kolorowego phi Kolersa, zobaczysz, jak zupełnie inne zjawiska zostają wyjaśnione za pomocą modelu wielokrotnych szkiców.

(5) Wiele powszechnie znanych eksperymentów, którym będziemy się przyglądać, jest uznawanych przez *niektórych* wybitnych badaczy za zaprzeczenie pewnego rodzaju konserwatywnej teorii materialistycznej, którą tu prezentuję, więc jeśli ma istnieć *naukowe* wyzwanie dla mojego wyjaśnienia świadomości, jest to pole walki, które zostało wybrane przez opozycję.

(6) Co więcej, omawiane zjawiska są fascynujące i warto je dobrze poznać[40].

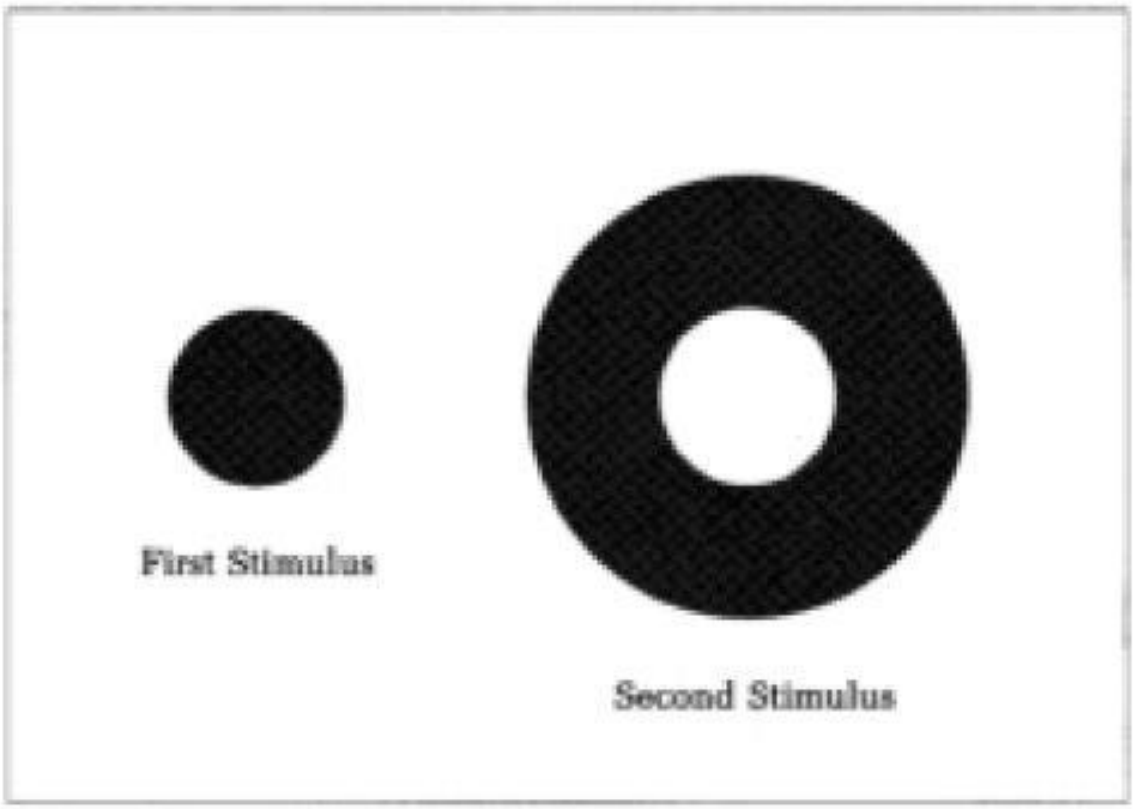
1. Ulotne momenty i skaczące króliki

Zwykle wystarczającym, choć niekoniecznym warunkiem przeżywania czegoś jest następujący po przeżyciu werbalny raport, i jest to podstawowa kwestia we wszystkich tych zagadkowych zjawiskach. Załóżmy, że mimo iż twój mózg zarejestrował – odpowiedział na – pewne aspekty jakiegoś wydarzenia, coś wkracza między tę wewnętrzną odpowiedź a następującą po nim możliwość werbalnego raportu. Jeśli nie było czasu czy okazji do początkowej jawnej odpowiedzi żadnego rodzaju i jeśli zdarzenia rozgrywające się w międzyczasie zapobiegają temu, aby późniejsze jawne odpowiedzi (werbalne lub inne) mogły dołączyć jakieś odniesienie do jakichś aspektów pierwszego wydarzenia, powstaje intrygujące pytanie: czy nigdy nie zostały one świadomie dostrzeżone, czy zostały szybko zapomniane?

W wielu eksperymentach zmierzono „zakres uwagi". W teście zakresu pamięci akustycznej słyszy się nagranie wielu szybko przedstawionych (powiedzmy czterech na sekundę), niezwiązanych ze sobą elementów, a osoba badana jest proszona o ich identyfikację. Nie jest w stanie odpowiedzieć, dopóki zdarzenie słuchowe się nie skończy, po czym identyfikuje niektóre z nich, a inne nie. Jednak subiektywnie słyszy je wszystkie wyraźnie i równie dobrze. Naturalnym pytaniem jest: czego dokładnie była świadoma? Nie ma wątpliwości, że wszystkie informacje z nagrania zostały przetworzone przez jej układ słuchowy, jednak czy cechy szczególne elementów, które później nie zostały nazwane, dostają się do jej świadomości, czy też były zarejestrowane nieświadomie? *Wydaje się*, że były w świadomości, ale czy *naprawdę*?

W innym paradygmacie eksperymentalnym zostaje pokazany slajd, na którym wydrukowano wiele liter. (Używa się tu *tachistoskopu*, urządzenia wyświetlającego, które można tak ustawić, aby pokazywało bodziec o ściśle określonej jasności na odpowiednią ilość milisekund – czasem tylko na 5 ms, czasem na 500 ms lub dłużej). Po obejrzeniu slajdu można zrelacjonować jedynie niektóre litery, ale reszta z pewnością była dostrzeżona przez osoby badane. Twierdzą one, że litery tam były, wiedzą, ile ich dokładnie było, i mają odczucie, że były bardzo wyraźne. Jednak nie potrafią ich wskazać. Czy bardzo szybko o nich zapomniały, czy tak naprawdę w ogóle nie zostały one przez nie świadomie dostrzeżone?

Głęboko studiowane zjawisko *metakontrastu* (Fehrer i Raab 1962) wyraźnie pokazuje sedno modelu wielokrotnych szkiców. (Analizę podobnego zjawiska znajdziesz u Breitmeyera 1984). Jeśli bodziec jest na krótką chwilę pokazany na ekranie (powiedzmy na 30 ms – mniej więcej na tak długo pokazywana jest klatka telewizyjna), a następnie natychmiast pokazany jest drugi bodziec „maskujący", osoby badane *relacjonują*, że widziały tylko drugi bodziec. Pierwszym bodźcem mógł być kolorowy krążek, a drugim kolorowy pierścień, którego wewnętrzna część jest blisko dopasowana do miejsca, gdzie wcześniej pojawił się dysk.



Ryc. 6.1

Na miejscu osoby badanej zobaczylibyśmy to na własne oczy; bylibyśmy gotowi przysięgać, że był tylko jeden bodziec: pierścień. W literaturze psychologicznej standardowy opis tego zjawiska jest stalinowski: drugi bodziec w jakiś sposób *zapobiega świadomemu* przeżywaniu pierwszego bodźca. Innymi słowy, w jakiś sposób zatrzymuje pierwszy bodziec w drodze do świadomości. Mimo to ludzie potrafią częściej niż przypadkowo stwierdzić, czy pojawił się jeden, czy dwa bodźce. Pokazuje nam to po raz kolejny, powie teoretyk stalinowski, że bodźce mogą wywierać na nas swój wpływ bez naszej świadomości. Pierwszy bodziec nigdy nie pokazuje się na scenie świadomości, jednak jego efekty są nieświadome. Możemy odpowiedzieć na to wyjaśnienie metakontrastu konkurencyjną hipotezą orwellowską: osoby badane są w rzeczywistości świadome pierwszego bodźca (co wyjaśniałoby ich umiejętność prawidłowego odgadnięcia liczby bodźców), ale ich wspomnienie tego świadomego doświadczenia zostaje *prawie* całkowicie wymazane z pamięci przez drugi bodziec (dlatego też zaprzeczają, że widziały pierwszy bodziec, pomimo znamiennego faktu, iż wyniki zgadywania są zbyt często poprawne, by były tylko wynikiem przypadku). Rezultatem jest impas – oraz zawstydzenie po obu stronach, gdyż żadna ze stron nie jest w stanie wskazać żadnego istotnego rezultatu eksperymentalnego, który rozwiązałby konflikt.

Model wielokrotnych szkiców radzi sobie z metakontrastem w następujący sposób. Gdy wiele zdarzeń pojawia się w krótkim czasie, mózg czyni upraszczające założenia. Zewnętrzny kontur krążka bardzo szybko zmienia się w wewnętrzny kontur pierścienia. Mózg, początkowo poinformowany jedynie o tym, że coś się wydarzyło (coś z okrągłym konturem w konkretnym miejscu), szybko otrzymuje potwierdzenie, że był to pierścień z wewnętrznymi i zewnętrznymi

konturami. Bez dalszych dowodów na to, że był to krążek, mózg dochodzi do konserwatywnego wniosku, iż pojawił się tylko pierścień. Czy powinniśmy twierdzić, że krążek był świadomie przeżywany, bo *gdyby nie przeszkodził pierścień*, krążek zostałby zrelacjonowany? Byłoby to błędem polegającym na założeniu, że moglibyśmy „zamrozić klatkę" w filmie w teatrze kartezjańskim i upewnić się, że zarys krążka rzeczywiście dostał się do teatru, zanim jego wspomnienie zostało wymazane z pamięci przez kolejne zdarzenie. Model wielokrotnych szkiców zakłada, że informacja o dysku przez chwilę odgrywała funkcjonalną rolę, która mogła doprowadzić do jej późniejszego zrelacjonowania, jednak ten stan się skończył; nie ma powodu, aby twierdzić, że stan ten był w zaklętym kręgu świadomości, zanim nie został zastąpiony czymś innym, ani przeciwnie – twierdzić, że nigdy tak naprawdę nie osiągnął tego uprzywilejowanego stanu. Szkice, które zostały stworzone w danych momentach i miejscach w mózgu, były później wycofane z obiegu, zastąpione wersjami zrewidowanymi, żaden z nich jednak nie może być wyróżniony jako definiujący treść *świadomości*.

Jeszcze bardziej zaskakującym pokazem takich możliwości rewizji jest *królik skórny*. Psychologowie Frank Geldard i Carl Sherrick opisali pierwotne eksperymenty z 1972 roku (zob. również Geldard 1977; Geldard i Sherrick 1983, 1986). Ręka osoby badanej leży zamortyzowana na stole, a mechaniczne młoteczki umieszczone są w dwóch lub trzech miejscach wzdłuż ręki, w odległości nie większej niż 30 centymetrów. Serie rytmicznych uderzeń przekazują młoteczki, na przykład pięć w nadgarstku, następnie dwa w okolicy łokcia, po czym trzy kolejne na ramieniu. Uderzenia są dostarczane z przerwami między bodźcami o długości od 50 ms do 200 ms. Tak więc ciąg uderzeń może trwać mniej niż sekundę lub nawet dwie czy trzy sekundy. Niesamowity efekt jest taki, że uderzenia zdają się osobie badanej poruszać w regularnych sekwencjach, w równych od siebie odległościach w górę ręki – jak gdyby małe zwierzątko skakało wzdłuż ich górnej kończyny. Z początku mamy ochotę zapytać: *jak mózg wiedział*, że po pięciu uderzeniach w nadgarstek nastąpią uderzenia w okolicy łokcia? Osoby badane są świadome „przejścia" uderzeń z nadgarstka, zaczynając od drugiego uderzenia, jednak w innych próbach, w których późniejsze uderzenia w łokieć nie następują, osoby badane czują wszystkie pięć uderzeń w nadgarstku właśnie tak, jak można by się spodziewać. Mózg oczywiście nie może „wiedzieć" o uderzeniu w łokieć, zanim ono się wydarzy. Jeśli wciąż urzeka cię teatr kartezjański, możesz chcieć spekulować, że mózg opóźnia świadome przeżycie do momentu, w którym wszystkie uderzenia zostają „dostarczone" do jakiejś stacji między ręką a ośrodkiem świadomości (gdziekolwiek miałoby to być), a w ten sposób stacja redaguje dane, aby dopasować je do teorii ruchu, i przesyła tak przygotowaną wersję do świadomości. Ale czy mózg zawsze opóźnia reakcję na jedno uderzenie na wypadek, gdyby pojawiły się kolejne? Jeśli nie, to skąd „wie", kiedy opóźniać?

Model wielokrotnych szkiców pokazuje, że to pytanie jest chybione. Przesunięcie w przestrzeni (wzdłuż ręki) jest rozróżniane w czasie przez mózg. Liczba uderzeń również jest rozróżniona. W rzeczywistości fizycznej uderzenia wystąpiły w konkretnych miejscach, jednak upraszczające założenie jest takie, że były w stałych odstępach w czasoprzestrzennym przebiegu przeżycia. Mózg przyjmuje tę skromną, choć błędną interpretację, oczywiście *po zarejestrowaniu* uderzeń, a w rezultacie wcześniejsze (częściowe) interpretacje uderzeń zostają wymazane, lecz ich skutki uboczne mogą nadal trwać. Na przykład załóżmy, że prosimy osoby badane, aby wcisnęły przycisk, gdy będą czuły *dwa uderzenia w nadgarstek*; nie byłoby zaskoczeniem, gdyby mogły rozpocząć wciskanie, *zanim* zostałyby rozróżnione uderzenia na przedramieniu i nie zinterpretowałyby drugiego uderzenia źle jako pojawiającego się gdzieś na przedramieniu.

Musimy wystrzegać się błędu, jakim jest założenie, że treść, którą wywnioskowalibyśmy z tak wczesnego badania, składałaby się na „pierwszy rozdział" treści, którą znaleźlibyśmy

w narracji, gdybyśmy zbadali to samo zjawisko w późniejszym momencie. Jest to mylenie dwóch różnych „przestrzeni": przestrzeni reprezentowania i reprezentowanej przestrzeni. Jest to tak kusząca i typowa pomyłka, że zasługuje na swój własny podrozdział.

2. Jak mózg reprezentuje czas

Materializm kartezjański – pogląd, którego nikt nie jest zwolennikiem, jednak narzuca się prawie każdemu – sugeruje następujący obraz podświadomości. Wiemy, że informacje wędrują w mózgu i są przetwarzane przez przeróżne mechanizmy w rozmaitych miejscach. Intuicja sugeruje nam, że nasze strumienie świadomości składają się z sekwencyjnych zdarzeń oraz że w każdym momencie każdy z elementów tej sekwencji może zostać określony jako ten, który już pojawił się „w świadomości", lub taki, który się jeszcze „tam" nie pojawił. Jeśli rzeczywiście tak jest, wydaje się, że nośniki reprezentacji wędrujące w mózgu muszą być jak wagony na torach; kolejność, w jakiej przekraczają pewien punkt, będzie kolejnością, z jaką „przybędą" do teatru świadomości i (w ten sposób) staną się świadome. Aby stwierdzić, *gdzie* w mózgu umiejscowiona jest świadomość, należy prześledzić wszystkie tory nośników informacji i zobaczyć, który punkt mijają konkretne nośniki w momencie, gdy ich treść staje się świadoma.

Po krótkim namyśle nad podstawowym zadaniem mózgu zauważymy, co z tym obrazem jest nie tak. Zadaniem mózgu jest kierowanie ciałem w zmiennym świecie, pełnym nagłych niespodzianek, zatem musi on zbierać informacje ze świata i korzystać z nich *szybko*, aby „wytworzyć przyszłość" – przewidywać, aby pozostać na jeden krok przed katastrofą (Dennett 1984a, 1991b). Mózg musi więc reprezentować czasowe właściwości zdarzeń w świecie i musi to robić sprawnie. Procesy odpowiedzialne za wykonywanie tego zadania są rozmieszczone przestrzennie w dużym mózgu bez centrali, a komunikacja między obszarami jest stosunkowo wolna; elektrochemiczne impulsy nerwowe podróżują tysiące razy wolniej niż światło (lub sygnały elektroniczne w przewodach). Mózg jest zatem pod istotną presją czasu. Często musi wprowadzać zmiany w docierających do ciała informacjach w świetle nowych danych w okienku czasowym, które nie pozwala na opóźnienia. Po stronie wejścia istnieją zadania związane z analizą percepcyjną, jak w przypadku percepcji mowy, które przekraczałyby fizyczne ograniczenia maszynerii mózgowej, gdyby mózg nie stosował pewnych genialnych strategii przewidywania, opierających się na nadmiarowości danych wejściowych. Normalna mowa odbywa się w tempie czterech lub pięciu sylab na sekundę, jednak urządzenia analizujące gramatykę w naszym mózgu wyewoluowały na tak potężne, że mogą poddać analizie „skompresowaną mowę" – w której słowa są elektronicznie przyspieszone bez podnoszenia tonu w stylu wiewiórki – o tempie do 30 sylab na sekundę. Jeśli natomiast chodzi o wyjście, wiele rzeczy musi się wydarzyć tak szybko i precyzyjnie, że mózg nie ma czasu dostosować swoich sygnałów sterujących w świetle otrzymywanej informacji zwrotnej; czynności takie jak gra na pianinie czy precyzyjny rzut kamieniem (Calvin 1983, 1986) muszą być zainicjowane *balistycznie*. (Czynności balistyczne przypominają niesterowane pociski; gdy zostaną wystrzelone, ich tor nie może być zmieniony).

W jaki zatem sposób mózg na bieżąco śledzi potrzebne mu informacje na temat czasu? Weźmy następujący problem: odległość od palca u nogi do mózgu jest większa niż z biodra do mózgu, ramienia do mózgu czy z czoła do mózgu, więc bodźce otrzymane w tym samym momencie w tych różnych miejscach dotrą do centrali w różnym czasie, przy założeniu, że prędkość transmisji informacji jest stała na wszystkich ścieżkach. Można zapytać: jak mózg „zapewnia centralną jednoczesność reprezentacji dla jednoczesnych bodźców dystalnych"? Angażując się w pewnego rodzaju spekulatywną, odwrotną inżynierię, moglibyśmy pomyśleć:

być może wszystkie aferentne drogi nerwowe są jak zwijana taśma miernicza – wszystkie tej samej długości: nerwy prowadzące do palców u stóp są całkowicie odwinięte, te prowadzące do czoła są natomiast zwinięte w mózgu. Sygnały z tych drugich podążają po zawiniętym, opóźniającym torze, docierając do centrali dokładnie w tym samym momencie, w którym docierają do niej sygnały z palca, poruszające się po niezwiniętych torach. Można by sobie również wyobrazić, że dłuższe drogi nerwowe mają zwężoną średnicę (niczym makaron domowej roboty) oraz że prędkość transmisji zależy od średnicy. (Tak się rzeczywiście dzieje, ale niestety w odwrotnym kierunku! Grubsze włókna prowadzą impulsy szybciej). Są to sugestywne (i głupkowate) modele mechanizmów, które rozwiązałyby nasz problem, jednak pierwszym błędem jest myślenie, że mózg w ogóle musi rozwiązywać ten problem. Mózg nie powinien go rozwiązywać z oczywistego powodu inżynieryjnego: trwoni cenny czas, planując swoje działania na podstawie harmonogramu dostosowanego do „najgorszego przypadku". Dlaczego bardzo ważne informacje z (na przykład) czoła miałyby czekać w przedpokoju tylko dlatego, że być może pewnego dnia nadarzy się okazja, że równoczesne sygnały z placów u nóg będą musiały jakoś z nimi się zbiec?[41]

To funkcjonowanie komputerów cyfrowych opiera się na takich opóźnieniach, aby przygotować się na najgorsze oraz aby zapewnić synchronizację. Mechanizm działający w sumatorze równoległym, który powstrzymuje sumowanie do czasu, aż zostanie ono wyzwolone przez impuls synchronizacji, przypomina pozwijane nerwy. Konstruktorzy superkomputerów muszą zaś zapewnić ściśle identyczną długość przewodów łączących różne części, co czasem wymaga użycia dodatkowych pętli przewodów. Komputery mogą sobie jednak pozwolić na taką miejscową niewydajność, gdyż i tak szybko działają. (Zasadność tych niewielkich czasowych spowolnień staje się coraz bardziej wątpliwa w związku z rynkiem wymagającym coraz szybszych komputerów; wiele z nich jednak funkcjonuje nadal głównie dlatego, że inżynierowie nie wiedzą, jak zaprojektować całkowicie asynchroniczny system komputerowy, nieregulowany żadnym głównym układem synchronizacji).

Odgórna regulacja synchroniczności wymaga opóźnień. Parając się inżynierią odwrotną, możemy spekulować, że jeśli istnieją w mózgu wydajne sposoby reprezentacji informacji o czasie, które unikają tych opóźnień, ewolucja z pewnością by je „odnalazła". Okazuje się, że istnieją takie sposoby, które możemy zilustrować pewnym incydentem historycznym, ukazującym to zjawisko w ogromnym powiększeniu – tak w czasie, jak i w przestrzeni.

Warto sobie uzmysłowić trudności komunikacyjne, jakie musiało mieć rozległe Imperium Brytyjskie, zanim pojawiło się radio czy telegraf. Kontrolowanie światowego imperium z centrali w Londynie nie zawsze było wykonalne. Najbardziej znany incydent to z pewnością bitwa pod Nowym Orleanem, 2 stycznia 1815 roku, piętnaście dni po podpisaniu w Belgii rozejmu kończącego wojnę roku 1812. W tej bezsensownej walce zostało zabitych ponad tysiąc brytyjskich żołnierzy. Owa klęska unaoczni, jak działał ten system. Załóżmy, że pierwszego dnia zostaje podpisany układ w Belgii, a informacje o nim zostają rozesłane drogą lądową i wodną do Ameryki, Indii czy Afryki. Piętnastego dnia rozgrywa się bitwa pod Nowym Orleanem, a informacje o porażce zostają rozesłane do Anglii, Indii itd. Dwudziestego dnia, niestety za późno, do Nowego Orleanu dociera informacja o traktacie (wraz z rozkazem poddania się). Załóżmy, że trzydziestego piątego dnia wiadomość o porażce dochodzi do Kalkuty, ale informacja o traktacie dociera dopiero czterdziestego dnia (gdyż została przesłana powolną drogą lądową). Gubernatorowi generalnemu Indii „wydawałoby się", że bitwa odbyła się przed podpisaniem traktatu – gdyby nie zwyczaj datowania listów, który pozwala mu zrozumieć całą sytuację[42].

Te oddalone od siebie ośrodki rozwiązywały większość problemów przekazywania

informacji dotyczących czasu, osadzając reprezentacje tych informacji w *treści* owych sygnałów, dzięki czemu czas przybycia sygnału był *nieistotny* dla informacji, którą w sobie niósł. Data napisana na początku listu (lub stempel z datą na kopercie) przekazuje odbiorcy informację o tym, kiedy sygnał został nadany, informację, która pozostaje niezmieniona mimo wszelkich opóźnień w dostarczeniu[43]. Rozróżnienie reprezentowanego (przez stempel) czasu i czasu reprezentowania (dzień, w którym list dociera) jest przykładem znanego rozróżnienia treści i nośnika. To konkretne rozwiązanie nie jest dostępne dla komunikujących się między sobą obszarów mózgu (gdyż „nie znają daty" wysłania danych), ogólna zasada rozróżnienia treść/nośnik ma głębszy związek z modelami mózgu w kategoriach przetwarzania informacji, niż zwykle się uważa[44].

Ogólnie rzecz biorąc, musimy odróżnić własności tego, co reprezentuje, od własności tego, co reprezentowane. Ktoś mógłby z całych sił krzyknąć: „po cichu, na paluszkach!", istnieją ogromne obrazki mikroskopijnych obiektów i nie jest niemożliwe namalowanie artysty szkicującego węglem. Pierwsze zdanie opisu stojącego mężczyzny nie musi dotyczyć jego głowy, a ostatnie jego stóp. Zasada ta odnosi się również, w sposób mniej oczywisty, do czasu. Zwróćmy uwagę na *wypowiedzianą frazę* „jasny, krótki błysk czerwonego światła". Jej początek to „jasny", a koniec to „czerwonego światła". Te fragmenty wypowiedzi nie reprezentują same w sobie początku i końca krótkiego czerwonego błysku (podobnie argumentuje Efron 1967, s. 714). Żadne zdarzenie w układzie nerwowym nie może mieć zerowego czasu trwania (tak samo jak nie może mieć zerowej rozciągłości przestrzennej), ma zatem początek i koniec oddzielone od siebie jakimś wycinkiem czasu. Jeśli to zdarzenie samo w sobie *reprezentuje* zdarzenie w przeżyciu, to zdarzenie reprezentowane też nie ma zerowego czasu trwania; ma więc początek, środek i koniec. Nie ma jednak żadnego powodu, aby twierdzić, że początek reprezentowania jest początkiem tego, co jest reprezentowane[45]. W rzeczywistości różne cechy są analizowane przez różne neuronalne funkcje w różnym tempie (np. lokalizacja, kształt i kolor), a gdyby poproszono nas o reakcję na każdą z tych cech osobno, zrobilibyśmy to z różnym opóźnieniem, choć postrzegamy wydarzenia, a nie sukcesywnie analizowany ciąg elementów czy cech percepcyjnych[46].

Powieść obyczajowa lub historyczna nie musi być stworzona w kolejności, jaką ma koniec końców przedstawić – czasem autorzy zaczynają od końca i wracają do początku. Co więcej, taka opowieść może zawierać retrospekcje, w których wydarzenia są *reprezentowane, jakby* wydarzyły się w pewnej kolejności, za pomocą *nośników reprezentujących*, występujących w innej kolejności. Podobnie reprezentowanie w mózgu *A przed B* nie musi być uzyskane poprzez:

najpierw reprezentację A,

a następnie reprezentację B.

Stwierdzenie „B po A" jest przykładem (mówionego) nośnika, który reprezentuje A będące przed B, a mózg może wykorzystać tę samą wolność umiejscawiania w czasie. Dla mózgu jest ważne niekoniecznie to, kiedy poszczególne zdarzenia reprezentowania występują w różnych częściach mózgu (dopóki następują na tyle szybko, że mogą sterować tym, co musi być sterowane!), ale ich *treść czasowa*. Innymi słowy, ważne jest dalsze sterowanie zdarzeniami przez mózg „przy założeniu, że A wystąpiło przed B" bez względu na to, czy informacja o tym, że A nastąpiło, dociera do odpowiedniego systemu w mózgu i zostaje rozpoznana przed informacją, że nastąpiło B, czy też po niej. (Przypomnijmy sobie dowódcę w Kalkucie: najpierw został poinformowany o bitwie, a następnie o rozejmie, jednak dostał również informację o tym, że rozejm był jako pierwszy, więc mógł zachować się odpowiednio do sytuacji. Musiał *ocenić*, że rozejm był przed bitwą; poza tym nie musiał organizować żadnej „historycznej rekonstrukcji",

w której otrzymałby listy we „właściwej” kolejności).

Niektórzy jednak uważają, że czas jako jedyny musi być reprezentowany przez mózg lub umysł *musi* być reprezentowany „przez samego siebie”. Filozof Hugh Mellor w swojej książce *Real Time* wyraża tę tezę jasno i stanowczo:

Załóżmy, że jestem świadkiem zdarzenia *e*, które występuje przed kolejnym zdarzeniem, *e**. Muszę najpierw zobaczyć *e*, a potem *e**, a obserwowanie *e* zostaje w jakiś sposób przypomniane podczas obserwowania *e**. To znaczy, że moje obserwowanie *e* ma wpływ na moje obserwowanie *e**: to właśnie sprawia, że widzę – trafnie bądź nie – *e* przed *e**, a nie odwrotnie. Jednak zobaczenie *e* przed *e** oznacza widzenie *e* jako pierwszego. W związku z tym przyczynowa kolejność mojego postrzegania tych zdarzeń, przez ustalenie czasowej kolejności, w jakiej je zaobserwowałem, ustala czasową kolejność spostrzeżeń samych w sobie [...]. Należy zwrócić uwagę [...] na szokujący fakt, że postrzeganie czasowego porządku potrzebuje czasowo uporządkowanych spostrzeżeń. *A zatem żadna inna cecha czy relacja nie musi być włączona w ich postrzeganie*: na przykład postrzeganie kształtu i koloru nie musi mieć odpowiedniego kształtu czy koloru. [Mellor 1981, s. 8, podkr. moje – D.C.D.]

Nie jest to prawda, jednak jest w tym fragmencie coś trafnego. Jako że podstawową funkcją reprezentacji w mózgu jest sterowanie zachowaniem w czasie rzeczywistym, czasowa synchronizacja elementów reprezentujących jest *do pewnego stopnia* istotą ich zadania, a to z dwóch względów.

Po pierwsze, na początkowym etapie procesu percepcyjnego umiejscowienie w czasie może być tym, co *wyznacza treść*. Zwróćmy uwagę, jak rozróżniamy punkty, z których jeden porusza się na ekranie kinowym ze strony prawej na lewą, a drugi z lewej na prawą. *Jedyną* różnicą między nimi może być kolejność, w jakiej (co najmniej) dwie klatki są wyświetlane na ekranie. Jeśli najpierw wyświetla się A, a potem B, punkt jest widoczny jako poruszający się w jednym kierunku; jeśli najpierw wyświetla się B, a potem A, punkt jest widziany jako poruszający się w odwrotnym kierunku. Jedyną różnicą między bodźcami, którą mógłby wykorzystać mózg, aby rozróżnić ten kierunek, jest kolejność, w jakiej się one pojawiają. To rozróżnienie jest zatem, ze względów logicznych, oparte na szczególnie wyraźnym rozróżnianiu kolejności czasowej przez mózg. Ponieważ klatki filmowe są zazwyczaj przedstawiane z prędkością 24 na sekundę, wiemy, że układ wzrokowy potrafi ustalić kolejność bodźców, które występują co około 50 ms. To oznacza, że rzeczywiste czasowe właściwości sygnałów – czas ich rozpoczęcia, ich prędkość w układzie, a więc też ich czas dotarcia – muszą być ściśle kontrolowane, dopóki takie rozróżnienie nie nastąpi. W przeciwnym razie informacje, na których to rozróżnienie musi się opierać, zginą bądź zostaną ukryte.

Zjawisko to możemy zaobserwować na większą skalę na początku regat żeglarskich; *widzimy* żaglówkę przecinającą linię startu, a następnie *słyszymy* wystrzał z pistoletu startowego, ale czy żaglówka wystartowała zbyt wcześnie? Stwierdzenie tego jest logicznie niemożliwe, dopóki nie przeliczymy prędkości rozchodzenia się dźwięku i światła do miejsca, z którego obserwowaliśmy start. Gdy już ustalimy, jak naprawdę wyglądała sytuacja (*wszystko w porządku* lub *falstart żaglówki nr 7*), treść może zostać przekazana uczestnikom bez pośpiechu, bez względu na to, jak szybko czy daleko wiadomość musi się przemieścić.

Zatem umiejscowienie w czasie elementu reprezentującego jest istotne *do czasu* poczynienia rozróżnienia, na przykład *ze strony lewej na prawą* (lub *falstart*), lecz gdy to już się stanie w jakimś miejscu dzięki jakiemuś obwodowi w korze (lub obserwatorowi z komisji), treść tego sądu może zostać wysłana bez dbałości o kolejność w czasie do wszystkich miejsc w mózgu, które mogą tej informacji potrzebować. Tylko w ten sposób możemy wyjaśnić zastanawiający fakt, że ludzie mogą pomylić się w więcej niż połowie przypadków, oceniając

pewne kolejności czasowe, jednocześnie osądzając bezbłędnie inne sytuacje (na przykład dotyczące kierunku ruchu), które wymagają jeszcze większej precyzji czasowej. Korzystają z wyspecjalizowanych (i specjalnie umiejscowionych) elementów rozróżniających, aby wydawać bardzo trafne osądy.

Kolejne ograniczenie nałożone na umiejscawianie elementów w czasie zostało już wspomniane na marginesie: nie ma znaczenia, w jakiej kolejności są elementy reprezentujące, dopóki występują w czasie, który pozwala na odpowiednią regulację zachowania. Funkcja takiego elementu reprezentującego może zależeć od zdążenia przed *ostatecznym terminem*, co jest czasową własnością reprezentującego nośnika. Jest to szczególnie widoczne w sytuacjach presji czasowej, na przykład w planowanej swego czasu Inicjatywie Obrony Strategicznej[47]. Problemem nie jest to, jak komputery miałyby w trafny sposób reprezentować wystrzelenie pocisku, ale jak trafnie reprezentować wystrzelenie pocisku w tym krótkim czasie, w którym ciągle jeszcze można na coś wpłynąć. Wiadomość, że pocisk został wystrzelony o godzinie 6:04:23 678 wschodnioamerykańskiego czasu urzędowego, być może na zawsze będzie reprezentować czas wystrzelenia, jednak jej użyteczność może całkowicie zniknąć o godzinie 6:05 czasu wschodnioamerykańskiego. Zatem każde zadanie sterowania posiada *czasowe okno sterowania*, w którym parametry czasowe elementów reprezentujących mogą z zasady być przenoszone dowolnie.

Ostateczne terminy, które ograniczają takie okna, nie są sztywne, a raczej zależą od zadania. Jeśli zamiast przechwytywać pociski, spisujesz wspomnienia lub odpowiadasz na pytania podczas przesłuchań związanych z aferą Watergate (Neisser 1981), informacje dotyczące kolejności zdarzeń z twojego życia, które próbujesz sobie przypomnieć, aby regulować swoje działania, mogą być wyszukiwane w jakiejkolwiek kolejności i trzeba się spieszyć z wyciąganiem wniosków. Weźmy inny przypadek, bliższy zjawisku, które rozważamy. Wyobraźmy sobie, że dryfujemy w łódce i zastanawiamy się, czy zbliżamy się, czy też oddalamy od niebezpiecznej rafy, którą widać w oddali. Załóżmy, że *teraz* znamy odległość od rafy (na przykład mierząc kąt, któremu odpowiada ona w polu widzenia obserwatora); by odpowiedzieć na swoje pytanie, możemy poczekać chwilę i ponownie zmierzyć kąt, a jeśli pół godziny temu zrobiliśmy zdjęcie rafie, możemy zmierzyć ten sam kąt na zdjęciu, wykonać kilka obliczeń i wywnioskować, jak daleko byliśmy wówczas. Chcąc ocenić kierunek, w którym dryfujemy, musimy zmierzyć dwie odległości: na przykład odległość w południe i o 12:30, ale nie ma znaczenia, którą odległość obliczymy jako pierwszą. Lepiej jednak, żebyśmy liczyli na tyle szybko, aby zacząć wiosłować, zanim będzie za późno.

Reprezentacja czasu przez mózg jest zatem dwojako zakorzeniona w samym czasie: samo umiejscowienie w czasie nośnika reprezentującego może dostarczać świadectw lub wyznaczać treść, a cały sens reprezentowania czasu zdarzeń może się zagubić, jeśli nośnik reprezentujący nie pojawi się na tyle szybko, aby na cokolwiek wpłynąć. Sądzę, że Mellor docenia oba te czynniki i miał je na myśli, pisząc to, co wcześniej zacytowałem, jednak popełnia naturalny błąd, myśląc, że działając jednocześnie, *całkowicie* ograniczają reprezentowanie czasu, czyli że kolejność nośników reprezentujących *zawsze* pokazuje kolejność treści. Według niego nie zachodzi czasowe „zamazywanie", ja natomiast uważam, że *musi* ono występować – na małą skalę – *ponieważ* musi też występować zamazywanie przestrzenne (na małą skalę) punktu widzenia obserwatora.

Przyczyny muszą poprzedzać skutki. Ta fundamentalna zasada gwarantuje, że czasowe okna sterowania są ograniczone na obu końcach: z jednej strony przez najwcześniejszy moment, w którym informacja może dotrzeć do systemu, a z drugiej strony przez ostatni moment, w którym informacja mogłaby wpłynąć na sterowanie zachowaniem. Nie wiedzieliśmy jeszcze,

jak mózg może korzystać z czasu dostępnego w czasowym oknie sterowania, aby zanalizować informacje, które otrzymuje, i przekształcić je w spójną „narrację" służącą do kontrolowania reakcji.

Jak więc własności czasowe można wywnioskować z procesów mózgowych? Systemy „datowników" i „stempli pocztowych" teoretycznie nie są niemożliwe, istnieje jednak tańszy, mniej niezawodny, ale biologicznie bardziej prawdopodobny sposób: coś, co moglibyśmy nazwać *rozstrzyganiem wrażliwym na treść*. Pomocną analogią może tu być studio filmowe, w którym ścieżka dźwiękowa jest „synchronizowana" z filmem. Rozmaite segmenty nagrania mogły zgubić znaczniki czasu i nie istnieje żaden prosty sposób złożenia ich z odpowiednimi obrazami. Przesuwanie ich w przód i w tył w stosunku do obrazu i poszukiwanie zbieżności zwykle szybko da „najlepsze dopasowanie". Klaps filmowy na początku każdego ujęcia – „scena trzecia, ujęcie siódme, kamera, akcja!" – jest w dwójnasób wyrazisty, jako klaps dźwiękowy i wizualny, które można zsynchronizować, a wraz z nimi dalszą część nagrania. Zwykle jednak jest tyle wyrazistych punktów zgodności, że owa zgodność konwencjonalna na początku każdego ujęcia jest po prostu przydatną nadmiarowością. Otrzymanie właściwego nagrania zależy od *treści* obrazu i dźwięku, ale nie od ich głębokiej analizy. Dla montażysty nieznającego języka japońskiego synchronizacja japońskiego dźwięku i obrazu w filmie byłaby trudna i nudna, ale nie niemożliwa. Co więcej, kolejność etapów procesu składania kawałków w całość jest niezależna od treści produktu; montażysta może przygotować scenę trzecią, zanim zabierze się za drugą, i teoretycznie mógłby zacząć całą pracę od końca, potem wracając do początku.

Dosyć „głupi" proces może w mózgu podobnie przesuwać informacje, aby rozstrzygać o ich kolejności. Na przykład określanie głębokości stereogramów złożonych z przypadkowych punktów (Ryc. 5.7) jest problemem przestrzennym, do którego można łatwo porównać uzgadnianie czasowe. Ogólnie rzecz biorąc, mózg może rozwiązać część problemów związanych z czasowym wnioskowaniem właśnie dzięki temu procesowi, nie tyle wykorzystując dane z lewego i prawego oka, ile korzystając z dowolnych źródeł dostępnych dla procesu wymagającego oceny czasowej.

Wynikają z tego dwa ważne wnioski. Po pierwsze, tego rodzaju wnioskowanie czasowe może się opierać na (tego rodzaju czasowe rozróżnienia mogą być oparte na) porównaniach (niskopoziomowej) *treści* z kilku tablic danych, a ten proces działający w czasie rzeczywistym nie musi przebiegać w kolejności, którą ostatecznie będzie reprezentował jego wytwór. Po drugie, gdy taki czasowy wniosek zostanie wyciągnięty, co może stać się, *zanim* własności z poziomu wyższego zostaną uzyskane przez inne procesy, to nie musi być on wyciągany ponownie! Nie musi istnieć *późniejsza* reprezentacja, w której właściwości z wyższego poziomu są „prezentowane" po kolei w czasie rzeczywistym ku pożytkowi drugiego procesu oceniania kolejności. Innymi słowy, wyciągając wnioski z tego zestawienia czasowych informacji, mózg może przystąpić do reprezentowania rezultatów w formacie, który odpowiada jego potrzebom i środkom – a niekoniecznie w formacie, w którym „czas reprezentuje czas".

3. Przypadek „odesłania do przeszłości" Libeta

Określiliśmy sposób, w jaki mózg przetwarza informacje czasowe, ignorując umiejscowienie w czasie („czas przybycia") niektórych reprezentacji. Musimy jednak ponownie zwrócić uwagę na to, pod jaką presją czasową działa mózg. Patrząc na tę kwestię od końca, czyli od ostatecznego terminu, cała treść, zrelacjonowana bądź w inny sposób wyrażona w późniejszym zachowaniu, musi być obecna (w mózgu, ale niekoniecznie w „świadomości") na czas, aby móc wpłynąć na zachowanie. Na przykład, jeśli osoba badana w eksperymencie *mówi*

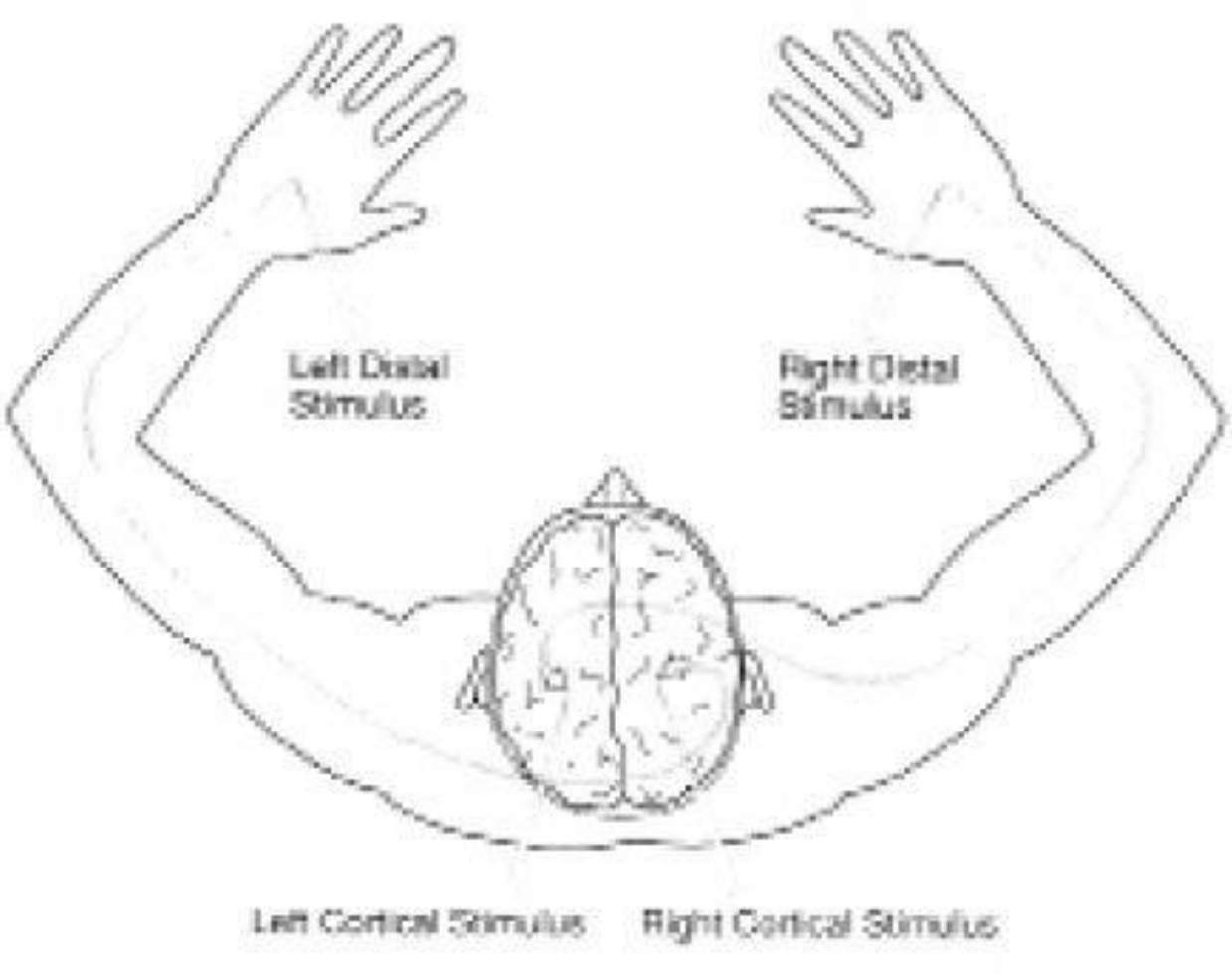
„pies” w odpowiedzi na bodziec wizualny, możemy cofnąć się od tego zachowania i stwierdzić, że z pewnością było ono sterowane procesem, który miał treść *pies* (chyba że osoba mówi „pies” w reakcji na każdy bodziec lub spędza cały dzień, powtarzając „pies, pies, pies...” itp.). Wiemy też, że potrzeba około 100 ms na *rozpoczęcie* wykonywania tego aktu mowy (i około 200 ms na jego skończenie), więc raczej możemy być pewni, że treść *pies* była obecna w obszarach mózgu odpowiedzialnych za język około 100 ms, zanim rozpoczęło się wypowiadanie słowa. Rozpoczynając od drugiej strony, możemy wskazać na najwcześniejszy moment, w którym treść *pies* mgła być obliczona lub wyodrębniona przez układ wzrokowy z danych wejściowych z siatkówki, a być może również prześledzić jej stworzenie i tor, po którym poruszała się do obszarów odpowiedzialnych za język.

Gdyby czas między bodźcem *pies* a wypowiedzeniem słowa „pies” okazał się krótszy niż czas fizycznie potrzebny do ustanowienia i przetransportowania przez system tej treści, byłaby to naprawdę anomalia (i z pewnością powód do utyskiwania i zgrzytania zębami). Jednak takie anomalie nie zostały odkryte. Odkryto natomiast pewne niesamowite zestawienia między dwoma ciągami przedstawionymi na rycinie 5.12. Kiedy próbujemy ustawić ciąg zdarzeń w obiektywnym strumieniu przetwarzania w mózgu w zgodzie z subiektywnym ciągiem osoby badanej, *wyznaczonym przez późniejszą wypowiedź osoby*, napotykamy czasem zaskakujące komplikacje. Taki przynajmniej jest wniosek, który moglibyśmy wyciągnąć z jednego z najbardziej dyskutowanych – oraz krytykowanych – eksperymentów w neuronauce: neurochirurgiczny eksperyment Banjamina Libeta ukazujący coś, co on sam nazywa „odesłaniem do przeszłości”.

Niekiedy podczas operacji ważne jest, aby pacjent był przytomny i uważny, znajdował się jedynie pod znieczuleniem miejscowym (to jak nowokaina u dentysty). Pozwala to chirurgowi na otrzymywanie natychmiastowej informacji zwrotnej od pacjenta na temat jego przeżyć, gdy badany jest jego mózg (zob. przypis 8 na stronie 84). Tę praktykę zapoczątkował Wilder Penfield (1958) i przez wiele lat neurochirurdzy zbierali dane dotyczące rezultatów bezpośrednich bodźców elektrycznych dostarczanych do różnych części kory. Od dawna wiadomo, że stymulacja obszarów w korze *somatosensorycznej* (pasek dogodnie ulokowany w poprzek górnej części mózgu) sprawia, iż pacjent doznaje wrażeń związanych z odpowiednimi częściami ciała. Na przykład stymulacja punktu znajdującego się w lewej części kory somatosensorycznej może spowodować uczucie krótkiego mrowienia w prawej dłoni osoby badanej (ze względu na znane odwrócenie w układzie nerwowym, które sprawia, że lewa półkula mózgu jest odpowiedzialna za prawą połowę ciała i *vice versa*). Libet porównał czas potrzebny na przeżycie mrowienia spowodowanego stymulacją kory oraz ten potrzebny na doznanie takiego samego wrażenia, ale bardziej tradycyjnie – dostarczając krótkiego impulsu elektrycznego do samej dłoni (Libet 1965, 1981, 1982, 1985b; Libet i inni 1979; zob. również Popper i Eccles 1977/1999; Dennett 1979b; Churchland 1981a, 1981b; Honderich 1984).

Czego moglibyśmy się spodziewać? Załóżmy, że dwaj pracownicy wyruszają do pracy codziennie o tej samej godzinie, ale jeden z nich mieszka na głębokich przedmieściach, a drugi parę ulic od biura. Jadą z tą samą prędkością, więc ze względu na dłuższą drogę, jaką musi pokonać pracownik z przedmieść, spodziewalibyśmy się, że dotrze on do biura później. Nie to jednak zaobserwował Libet, gdy zapytał pacjentów, co było pierwsze: mrowienie w dłoni wywołane pobudzeniem kory czy mrowienie w dłoni wywołane pobudzeniem samej dłoni. Opierając się na zebranych danych, stwierdził, że w obu przypadkach minęło sporo czasu (około 500 ms) między początkiem bodźca a „odpowiedniością neuronalną” (punkt, w którym – jak twierdzi Libet – następuje kulminacja procesów korowych i dochodzi do świadomego przeżycia mrowienia), lecz gdy stymulowana była dłoń, przeżycie było „automatycznie” „cofnięte

w czasie" i zostało odczute, jakby zdarzyło się, *zanim* pojawiło się mrowienie wywołane stymulacją mózgu.



Ryc. 6.2

Co więcej, Libet wspominał o przypadkach, w których lewa *kora* pacjenta była stymulowana, *zanim* nastąpiła stymulacja jego lewej *ręki*, co, jak można sądzić, doprowadziłoby do uczucia dwóch mrowień: najpierw w prawej ręce (mrowienie pochodzące od stymulacji kory), a następnie w ręce lewej. Jednak sprawozdanie osoby badanej było inne: „najpierw lewa, potem prawa".

Libet zinterpretował swoje rezultaty, silnie kwestionując materializm: „[...] różnica pomiędzy umiejscowieniem w czasie odpowiednich »umysłowych« i »fizycznych« zdarzeń wydaje się stanowić spore, aczkolwiek nie przytłaczające, trudności dla [...] teorii identyczności psychonerwowej (materialistycznej)" (Libet i inni 1979, s. 222). Według sir Johna Ecclesa, laureata Nagrody Nobla w dziedzinie medycyny za badania z zakresu neuropsychologii, tym trudnościom nie można podołać:

Ta procedura antydatowania wydaje się nie do wyjaśnienia na bazie procesów neurofizjologicznych. Przypuszczalnie jest to strategia, której nauczyć się może samoświadomy umysł [...] antydatowanie doznania zmysłowego można przypisać zdolności umysłu

samoświadomego do wykonania niewielkiego czasowego dopasowania, tj. pewnej sztuczki z czasem. [Popper i Eccles 1977/1999, t. 2, s. 182–183, przekład poprawiony]

Matematyk i fizyk Roger Penrose (1989/2000) uważa, że materialistyczne wyjaśnienie zjawiska Libeta wymagałoby rewolucji w fizyce. Eksperyment Libeta jest w kręgach pozanaukowych uważany za dowód prawdziwości dualizmu, jednak niewielu kognitywistów się z tym zgadza. Przede wszystkim ostro krytykuje się procedury eksperymentu Libeta oraz jego analizę rezultatów. Jego eksperyment nigdy nie został powtórzony, co dla wielu jest wystarczającym powodem, by w ogóle nie brać tych wyników pod uwagę. Sceptycy uważają więc, że takie zjawisko po prostu nie istnieje. *A gdyby istniało?* Jest to pytanie, które mógłby zadać filozof, lecz motywacja do jego zadania wykracza poza ramy filozoficzne. Nikt nie podważa istnienia mniej skomplikowanych zjawisk, takich jak phi czy „skórny królik", a ich interpretacja nastręcza tych samych problemów. Z teoretycznego punktu widzenia krótkowzroczne byłoby odrzucenie tego zjawiska ze względów *metodologicznych*, gdy pozostawia ono przypuszczenie, że gdyby eksperyment Libeta udało się kiedyś naprawdę powtórzyć, to byłyby to nader ponure wieści dla materializmu.

Pierwszą rzeczą, którą można zauważyć w eksperymencie Libeta, jest to, że nie dostarczyłby on żadnych dowodów na jakąkolwiek anomalię, gdybyśmy nie skorzystali z możliwości nagrania werbalnych relacji osób badanych na temat ich przeżyć, a następnie nie wykorzystali ich do stworzenia tekstu, a potem świata heterofenomenologicznego. Dźwięki, które wydają ze swoich strun głosowych podczas eksperymentu lub po nim, nie dają nawet cienia sprzeczności, jeśli traktowane są jedynie jako zjawiska akustyczne. Dźwięki nigdy nie są wydawane z głów, zanim poruszą się wargi, nie poruszają się też dłonie przed wydarzeniami mózgowymi, które je rzekomo powodują, wydarzenia nie następują również w korze, dopóki nie zostanie dostarczony bodziec, który jest ich źródłem. Obserwując i mierząc wydarzenia z eksperymentu, traktując je jedynie jako wewnętrzne i zewnętrzne zachowanie biologicznie zaimplementowanego układu sterowania ciałem, nie widzimy żadnego oczywistego naruszenia zwykłej, mechanicznej przyczynowości – której standardowy model zapewnia nam fizyka Galileusza i Newtona.

Można więc „pozbyć się problemu" i niczym bezkompromisowy behawiorysta odmówić poważnego potraktowania raportów introspekcyjnych. Nie jesteśmy jednak bezkompromisowymi behawiorystami; chcemy podjąć wyzwanie i zrozumieć to, co Libet nazywa „zasadniczo fenomenologicznym aspektem naszego ludzkiego istnienia w odniesieniu do funkcji mózgu" (1985a, s. 534). Libet niemalże rozumie heterofenomenologię. Mówi: „Ważne jest zdanie sobie sprawy, że te subiektywne cofnięcia i poprawki następują najwyraźniej na poziomie »sfery« *umysłowej*; nie są same w sobie widoczne w czynnościach na poziomie neuronów" (1982, s. 241). Skoro nie może *neutralnie* odwołać się do fenomenologii, musi tę anomalię przypisać „»sferze« umysłowej". Mały, choć wymuszony krok (*musi* go jednak uczynić, jeśli odrzuca behawioryzm) jest pierwszym krokiem prowadzącym wprost do dualizmu.

> Raporty osób badanych na temat ich przeżyć [...] nie były konstruktami teoretycznymi, lecz empirycznymi obserwacjami [...]. Metoda introspekcji ma swoje ograniczenia, jednak może być skutecznie wykorzystywana w ramach nauk przyrodniczych, a jest absolutnie niezbędna, jeśli staramy się zebrać eksperymentalne dane dotyczące problemu umysł-mózg. [Libet 1987, s. 185]

Raporty osób badanych, nawet przekształcone w tekst, są dla Libeta obserwacjami empirycznymi, ale to, *co relacjonują*, czyli zdarzenia w ich heterofenomenologicznych światach, to w rzeczywistości konstrukty teoretyczne. Libet twierdzi, że mogą być one użyte poprawnie w ramach nauk przyrodniczych, ale *tylko* jeśli od początku traktujemy je jako fikcje teoretyczne.

Libet głosi, że jego eksperymenty z bezpośrednią stymulacją kory pokazują „dwie

niesłychane własności czasowe":

(1) Zanim czynności mózgowe zapoczątkowane stymulacją sensoryczną *uzyskają „odpowiedniość neuronalną"* wystarczającą do wywołania świadomego przeżycia zmysłowego, *występuje istotne opóźnienie*.

(2) Po uzyskaniu odpowiedniości neuronalnej *subiektywne umiejscawianie zdarzeń z eksperymentu w czasie jest (automatycznie) cofnięte w czasie*, co dzieje się dzięki „sygnałowi umiejscowienia w czasie" w formie początkowej reakcji kory mózgowej na bodziec zmysłowy.

„Sygnał umiejscowienia w czasie" jest *pierwszą* erupcją czynności, która pojawia się w korze (pierwotny potencjał wywołany) i występuje od 10 do 20 ms po stymulacji dłoni. Libet twierdzi, że odesłanie do przeszłości zatrzymuje się na tym sygnale umiejscowienia w czasie.

Model Libeta jest stalinowski: po pierwotnym potencjale wywołanym zachodzą różnorodne procesy redakcyjne w korze, zanim nastąpi moment „odpowiedniości neuronalnej", kiedy to rzutowany jest ukończony film. W jaki sposób jest rzutowany? Tutaj poglądy Libeta wahają się między podejściem ekstremalnym i umiarkowanym (zob. także Honderich 1984):

(1) *Rzutowanie wstecz*: w jakiś sposób informacje są cofane w czasie do jakiegoś teatru kartezjańskiego, gdzie skończony film jest rzutowany wraz z pierwotnymi potencjałami wywołanymi. (Pierwotne potencjały wywołane, jako „sygnały umiejscowienia w czasie", służą raczej jako klaps w filmie, dokładnie pokazując rzutnikowi, jak daleko w przeszłość należy wysłać przeżycie).

(2) Odesłanie do przeszłości: film jest rzutowany w czasie rzeczywistym, jednak niesie ze sobą coś w rodzaju stempla, przypominając widzowi, że musi odbierać te zdarzenia jako zaistniałe wcześniej. (W tym przypadku pierwotne potencjały wywołane służą jako daty, które mogą być *reprezentowane* na ekranie kartezjańskim z tytułem „Wieczór przed bitwą pod Waterloo" lub „Nowy Jork, lato 1942").

Pojęcie używane przez Libeta to *odesłanie*, a broni on go, przypominając nam „od dawna rozpoznawane i akceptowane" zjawisko odesłania przestrzennego, o którym mówi w kolejnym fragmencie.

Subiektywne odesłanie do przeszłości to dziwna koncepcja i być może nie jest z początku łatwa do przyjęcia. Istnieje jednak ogromny precedens w postaci od dawna rozpoznawanej i akceptowanej koncepcji subiektywnego odesłania w wymiarze przestrzennym. Na przykład obraz przeżywany w reakcji na bodziec wzrokowy ma subiektywną konfigurację i lokalizację przestrzenną, które znacznie różnią się od przestrzennej konfiguracji i lokalizacji czynności neuronalnych, które wywołują („subiektywnie odesłany") obraz. [Libet 1981, s. 183; zob. również Libet i inni 1979, s. 221; Libet 1985b]

Libet wyciąga jednak wniosek, że czasowe przesunięcie rodzi problem dla materializmu (dla „teorii identyczności psychonerwowej": Libet i inni 1979, s. 222), zatem albo uważa, że przesunięcie przestrzenne rodzi ten sam problem, albo nie zrozumiał swojej własnej obrony. Jeśli odesłanie przestrzenne – fakt, że to, co widzimy, wydaje się być na zewnątrz, a nie wewnątrz naszych mózgów – stanowi problem dla materializmu, to dlaczego Libet wskazuje, że jego własna praca jest nowym, ważnym argumentem na rzecz dualizmu? Z pewnością fakt odesłania przestrzennego jest o wiele lepiej potwierdzony niż odesłania czasowe, których istnienie musiał tak pomysłowo zademonstrować. Wydaje się jednak, ze Libet ma radykalną (a w każdym razie

dosyć dziwaczną) wizję odesłania przestrzennego jako pewnego rodzaju „rzutowania”:

[...] istnieją eksperymentalne świadectwa na rzecz poglądu, że subiektywna czy umysłowa „sfera” mogłaby w rzeczywistości „wypełnić” przestrzenne i czasowe luki. Jakże inaczej można by na przykład ujmować wspomnianą już gigantyczną, *znaną* niezgodność między subiektywnym obrazem wizualnym a konfiguracją czynności nerwowych prowadzącą do przeżycia obrazu? [Libet 1981, s. 196]

Wygląda więc na to, że rzutnik, którego w mózgu nie odnalazł Smythies, w rzeczywistości chowa się w „sferze” umysłowej[48].

Jak zatem Libet ustalił te dwie niesamowite czasowe właściwości? „Odpowiedniość neuronalna”, która według Libeta wymaga nawet 500 ms aktywności korowej, jest wyznaczona przez to, jak późno po początkowej stymulacji bezpośrednia stymulacja korowa może *ingerować* w zrelacjonowane następnie świadome przeżycie. Po tym kluczowym przedziale czasu bezpośrednia stymulacja kory zostałaby zrelacjonowana przez osobę badaną jako *późniejsze* przeżycie. (Dociera zbyt późno do pokoju redakcyjnego, aby została włączona do „ostatecznego druku” pierwszego doświadczenia bodźca, pojawia się więc w kolejnym odcinku). Dane Libeta sugerują niezwykle zmienne okno redakcyjne: „Warunkujący bodziec korowy mógłby zostać rozpoczęty ponad 500 ms po dotknięciu dłoni i mimo wszystko zmodyfikować przeżycie, jednak w większości przypadków efekty działania wstecz nie były zaobserwowane, gdy przedziały S-C były dłuższe niż 200 ms” (1981, s. 185). Libet bardzo ostrożnie definiuje odpowiedniość neuronalną w kategoriach wpływu na niespieszny raport werbalny: „osobę badaną poproszono o relację w kilka sekund po dostarczeniu obu [...] bodźców” (1979, s. 195), i twierdzi, że „umiejscowienie w czasie subiektywnego przeżycia należy odróżnić od reakcji behawioralnej (takiej jak czas reakcji), która może nastąpić, zanim pojawi się świadomość [...]” (Libet i inni 1979, s. 193).

To zastrzeżenie pozwala mu bronić konkurencyjnej interpretacji danych Patricii Churchland. Jest ona pierwszą „neurofilozofką” (zobacz jej książkę z roku 1986: *Neurophilosophy: Toward a Unified Science of the Mind-Brain*). Gdy po raz pierwszy przeczytałem o wynikach Libeta (Popper i Eccles 1977), zachęciłem ją do zapoznania się z nimi, a ona ostro je skrytykowała (Churchland 1981a). Starała się podważyć pierwszą tezę Libeta, dotyczącą długiego czasu powstawania „odpowiedniości neuronalnej” wystarczającej dla świadomości, prosząc osoby badane o powiedzenie „już”, gdy tylko uświadomiły sobie bodziec skórny, podobny do zastosowanego przez Libeta. Średni czas reakcji u dziewięciu osób badanych wynosił 358 ms, co według niej dowodziło, że musiały one uzyskać odpowiedniość neuronalną po najwyżej 200 ms (resztę czasu wykorzystując na reakcję werbalną).

Odpowiedź Libeta jest stalinowska: reakcja werbalna – powiedzenie „już” – może zostać rozpoczęta nieświadomie. „Nie ma nic magicznego czy wyjątkowo pouczającego, gdy reakcja motoryczna to zwerbalizowane słowo »już«, a nie częściej spotykane wciskanie palcem przycisku. [...] Umiejętność wykrycia bodźca i reakcji nań czy też poddania się innemu wpływowi psychologicznemu z jego strony, bez jakiejkolwiek świadomości bodźca, jest powszechnie akceptowana” (Libet 1981, s. 187–188). Natomiast na zarzut: „Co w takim razie osoby badane przez Churchland myślały, że robią, jeśli nie mówią, tak jak zostały o to poproszone, że są świadome bodźca?” Libet mógłby dać standardową odpowiedź stalinowską: W końcu uświadomiły sobie bodziec, jednak do tego czasu ich relacja werbalna została zainicjowana[49].

Z tego powodu Libet odrzuca badania czasu reakcji, w tym te przeprowadzone przez Churchland, gdyż stosują „podstawowe kryterium przeżycia subiektywnego o wątpliwej ważności” (1981, s. 188). Uważa, że osobom badanym należy dać czas: „Relacja jest zdana bez

pośpiechu w kilka sekund po każdej próbie, co pozwala osobom badanym introspekcyjnie zbadać świadectwa" (s. 188). W jaki zatem sposób radzi sobie z przeciwną możliwością, że ta powolna reakcja daje orwellowskiemu rewizjoniście w mózgu mnóstwo czasu na zastąpienie *prawdziwych* wspomnień świadomości wspomnieniami *fałszywymi*?

Raport po próbie wymaga oczywiście, aby procesy pamięci krótkotrwałej oraz przypominania funkcjonowały, jednak nie sprawia to osobom badanym żadnej trudności, o ile nie mają zaburzeń tych procesów. [Libet 1981, s. 188]

Jest to założenie tezy, która ma być dopiero udowodniona w sporze z obrońcami podejścia orwellowskiego, gotowymi wyjaśnić różne efekty jako rezultaty *normalnego* zapominania czy halucynacji w przypominaniu, w wyniku czego pierwotne, realne zdarzenie zostaje zatarte w świadomości i zastąpione innymi wspomnieniami. Czy Libet pozwolił osobom na zbyt długie dywagacje, czy to Churchland była w gorącej wodzie kąpana? Jeśli Libet chce, aby jego sposób mierzenia czasu był *uprzywilejowany*, musi podjąć walkę z kontrargumentami.

Libet niemalże odmawia stwierdzenia, po której stronie leży racja: „Przyznaję, że raport na temat relatywnego umiejscowienia w czasie nie może, sam w sobie, stanowić wskaźnika czasu »bezwzględnego« (czasu zegarowego) przeżycia: jak stwierdziłem, nie istnieje metoda uzyskania takiego wskaźnika" (1981, s. 188). Powtarza tu swoje wcześniejsze stwierdzenie, że wydaje się „nie istnieć metoda, którą można by wyznaczyć bezwzględne umiejscowienie w czasie przeżycia subiektywnego" (Libet i inni 1979, s. 193). Umyka mu jednak możliwość, że wynika to z nieistnienia takiego momentu bezwzględnego czasu (zob. także Harnad 1989).

W swojej krytyce Churchland (1981a, 1981b) również popełnia błąd polegający na nieodróżnianiu reprezentowanego czasu od czasu reprezentowania:

Te dwie hipotezy różnią się zasadniczo jedynie tym, kiedy poszczególne wrażenia *zostały odczute*. [1981a, s. 177; podkr. moje – D.C.D.]

Nawet jeśli założymy, że wrażenia wynikające z jednoczesnego wrażenia skórnego i LM [*medial lemniscus* – wstęgi przyśrodkowej] są *odczuwane dokładnie w tym samym momencie*, opóźnienie w odpowiedniości neuronalnej dla bodźca skórnego może być po prostu artefaktem takiego układu. [1981b, s. 494; podkr. moje – D.C.D.]

Załóżmy, że zostały wyeliminowane wszystkie takie artefakty, a wrażenia *nadal* „odczuwane są dokładnie w tym samym czasie". Jak Churchland zinterpretowałaby taki niepożądany rezultat? Czy znaczyłoby to, że istnieje czas *t*, w którym zostaje odczuty bodziec 1 i 2 (perspektywa antymaterialistyczna), czy jedynie to, że bodźce 1 i 2 zostają odczute (przeżyte) jako równoczesne? Churchland nie odrzuca wniosku, że gdyby wyniki Libeta zostały potwierdzone, zdruzgotałyby (jak on sam czasem twierdzi) materializm. W innym jednak miejscu Churchland słusznie zauważa, że „iluzje czasowe mogą być intrygujące, ale nie ma powodu zakładać, że jest w nich coś nienaturalnego, a już na pewno nie istnieje nic, co różniłoby je od iluzji przestrzennych czy ruchowych i sprawiło, że jako jedyne mają źródło niefizyczne" (1981a, s. 179). Mogłoby się to okazać tylko wówczas, gdyby iluzje czasowe były zjawiskami, w których czas *został wypaczony*; jeśli te *wypaczenia* wydarzają się w „nieprawidłowych" momentach, czeka nas coś bardziej rewolucyjnego.

Cóż więc z eksperymentami Libeta ze stymulacją kory? Są one interesującymi, ale nierozstrzygającymi próbami ustalenia, *jak mózg reprezentuje kolejność czasową*. Pierwotne potencjały wywołane mogą w pewien sposób służyć jako punkty odniesienia dla neuronalnych reprezentacji czasu, choć Libet tego nie pokazał, jak słusznie zauważyła Churchland. Mózg może też nietrwale przechowywać reprezentacje czasowe. Nie reprezentujemy widzianych obiektów jako istniejących na siatkówce, a raczej w różnych odległościach w świecie zewnętrznym; dlaczego mózg nie miałby reprezentować zdarzeń również jako wydarzających się [obiektywnie],

gdy ma to najbardziej „ekologiczny" sens? Gdy wykonujemy czynności manualne, standardem powinien być „czas koniuszków palców"; gdy dyrygujemy orkiestrą, „czas ucha" może uchwycić nagranie. „Podstawowy czas korowy" mógłby być domyślnym standardem (niczym średni czas Greenwich dla Imperium Brytyjskiego); jest to jednak kwestia, którą należałoby zbadać dogłębniej.

Rozstrzygnięcie tego sporu utrudnia fakt, że zarówno osoba, która go wywołała, jak i jej krytyczka nie odróżnili konsekwentnie reprezentowanego czasu od czasu reprezentowania. Gadali jak dziad do obrazu, Libet zajął stanowisko stalinowskie, a Churchland orwellowskie, oboje zaś najwyraźniej zgadzali się, że rzeczywiście istnieje dokładny moment, w którym (w czasie „bezwzględnym", jak powiedziałby Libet) następuje świadome przeżycie[50].

4. Twierdzenie Libeta o subiektywnym opóźnieniu świadomości intencji

Koncepcja bezwzględnego umiejscowienia przeżycia w czasie jest wykorzystywana w późniejszych eksperymentach ze „świadomymi intencjami" Libeta. Próbował on w nich eksperymentalnie wyznaczyć bezwzględne umiejscowienie w czasie, pozwalając osobom badanym, które jako jedyne mają bezpośredni dostęp do swoich przeżyć, *zmierzyć czas samemu*. Prosił normalne osoby badane (a nie pacjentów neurochirurgicznych) o podjęcie „spontanicznych" decyzji zgięcia ręki w nadgarstku, jednocześnie zwracając uwagę na położenie wskazówki na tarczy (czyli wskazówki sekundowej zegara) dokładnie w momencie, w którym powzięli decyzję o zgięciu (Libet 1985a, 1987, 1989). Następnie (kilka sekund później) osoby te relacjonowały, gdzie znajdywała się wskazówka w momencie podjęcia decyzji. Pozwoliło to Libetowi wyliczyć, która była godzina (z dokładnością do milisekund), gdy osoby *myślały*, że powzięły decyzję, i porównać ten moment z umiejscowieniem w czasie zdarzeń zachodzących w ich mózgach. Znalazł świadectwa na rzecz tezy, że owe „świadome decyzje" pozostawały około 350–400 ms w tyle za początkiem „potencjałów gotowości", które uzyskał z elektrod umieszczonych na głowie. Te potencjały, jak twierdzi, wskazują na zdarzenia neuronalne determinujące dobrowolnie wykonywane czynności. Uważa, że „mózgowa inicjacja spontanicznej, dobrowolnej czynności rozpoczyna się nieświadomie" (Libet 1985a, s. 529).

W związku z tym wydawać by się mogło, że świadomość jest spóźniona w stosunku do procesów mózgowych, które faktycznie sterują ciałem. Wielu odbiera to jako niepokojącą, a nawet przygnębiającą perspektywę, gdyż wyklucza ona realną (a nie iluzoryczną) „rolę kierowniczą" „świadomego ja". (Zob. również dyskusje wielu komentatorów na temat artykułów Libet 1985a, 1987, 1989; oraz w Harnad 1982; Pagels 1988, s. 233ff; Calvon 1989a, s. 80–81).

Libet, jaśniej od większości krytyków, mówi o konieczności odróżnienia treści od nośnika: „Nie możemy mylić tego, *co* jest relacjonowane przez osobę badaną, z tym, *kiedy* może być ona introspekcyjnie świadoma tego, co relacjonuje" (1985a, s. 559). Poza tym stwierdza (s. 559), że osąd jednoczesności nie musi jednocześnie zapaść czy się pojawić; może dojrzewać przez jakiś czas (można to porównać do czasu potrzebnego sędziom na torze wyścigowym na analizę zdjęcia przekraczanej linii mety, na której to analizie opierają swoją ocenę, kto jest zwycięzcą).

Libet zebrał dane dotyczące dwóch szeregów czasowych:

(1) obiektywnego, składającego się z czasowych własności zewnętrznego zegara i istotnych czynności neuronalnych: potencjałów gotowości oraz elektromiogramów (EMG), które rejestrowały początek spinania mięśni;

(2) subiektywnego (zrelacjonowanego później), składającego się z obrazowania

umysłowego, wspomnień i planowania oraz z jednej informacji, będącej punktem odniesienia: osądu jednoczesności w postaci: *moja świadoma intencja (W) nastąpiła równocześnie z chwilą, w której wskazówka zegara znajdowała się w miejscu M*.

Wydaje się, że Libet chciał się zbliżyć do ulotnego *acte gratuit* dyskutowanego przez egzystencjonalistów (np. Gide 1948/1995; Sartre 1943/2007), wyboru zupełnie pozbawionego motywacji – i w związku z tym w pewnym sensie „wolnego". Jak zauważyło wielu komentatorów, tak niezwykłe czynności (które można by nazwać aktami celowej pseudoprzypadkowości) trudno uznać za paradygmatyczne przypadki „zwykłych czynów dobrowolnych" (Libet 1987, s. 784). Czy jednak w każdym razie w takim układzie eksperymentalnym wskazał on kategorię świadomych doświadczeń, jakkolwiek by ich nie scharakteryzować, którym można przyznać bezwzględne umiejscowienie w czasie?

Libet twierdzi, że gdy świadome intencje działania (przynajmniej pewnego rodzaju) zarejestruje się wraz ze zdarzeniami mózgowymi, które faktycznie inicjują te działania, widoczne jest opóźnienie tych pierwszych w granicach 300–500 ms. To bardzo dużo – nawet do pół sekundy – i z pewnością wygląda złowieszczo dla kogoś, kto jest przekonany, że nasze świadome akty *sterują* ruchami ciała. Wygląda to tak, jakbyśmy byli ulokowani w teatrze kartezjańskim, gdzie pokazuje się nam, z półsekundowym opóźnieniem, *prawdziwe* podejmowanie decyzji, które następuje *gdzie indziej* (gdzieś, gdzie nas nie ma). Nie jesteśmy „poza kręgiem wtajemniczonych" (jak mówią w Białym Domu), ale skoro nasz dostęp do informacji jest w ten sposób opóźniany, możemy co najwyżej w ostatniej chwili interweniować – myśląc „weto" lub „zgoda". Poniżej (nieświadomego) centrum dowodzenia nie mam żadnej prawdziwej inicjatywy, nie biorę udziału w narodzinach pomysłów, jednak mam w rękach odrobinę mocy kierowniczej i mogę wpływać na gotowe już taktyki działania przepływające przez mój gabinet.

Taka wizja robi wrażenie, ale jest niespójna. Model Libeta po raz kolejny jest stalinowski i nie istnieje żadna oczywista konkurencja orwellowska: badani byli świadomi swoich intencji wcześniej, lecz ta świadomość została wymazana z ich pamięci (bądź też zmieniona), zanim mieli możliwość przypomnienia sobie o niej. Libet przyznaje, że „jest to problem, jednak nie jest sprawdzalny eksperymentalnie" (1985a, s. 560).

Biorąc pod uwagę te słowa, czy zadanie ustalenia bezwzględnego umiejscowienia świadomości w czasie jest błędnie postawione? Libet ani jego krytycy nie doszli do tego wniosku. Libet, po odróżnieniu treści od nośnika – *co* jest reprezentowane od tego, *kiedy* jest reprezentowane – mimo wszystko z przesłanek dotyczących reprezentowanego przedmiotu próbuje wysnuć wniosek na temat bezwzględnego umiejscowienia nośników w czasie w świadomości. Psycholog Gerald Wasserman (1985, s. 556) wskazał ten problem: „Moment, w którym zewnętrzna, obiektywna wskazówka zajmuje dane miejsce na zegarze, może zostać ustalony bardzo łatwo, lecz nie jest to poszukiwany rezultat". Następnie wpada jednak w kartezjańską pułapkę: „Poszukiwanym rezultatem jest moment pojawienia się wewnętrznej, mózgowo-umysłowej reprezentacji wskazówki".

„Moment pojawienia się" wewnętrznej reprezentacji? Pojawienia się gdzie? Przede wszystkim istnieje nieprzerwana reprezentacja wskazówki (jej reprezentacja w różnych pozycjach) w różnorodnych obszarach mózgu, zaczynająca się na siatkówce i trwająca w całym układzie wzrokowym. Jasność wskazówki jest reprezentowana w jednych miejscach i momentach, jej lokalizacja w innych, a jej ruch w jeszcze innych. Wraz z poruszającą się zewnętrzną wskazówką wszystkie te reprezentacje zmieniają się w sposób asynchroniczny i rozproszony przestrzennie. Gdzie „wszystko to łączy się w momencie świadomości"? Nigdzie.

Wasserman trafnie zauważa, że zadanie osoby badanej, polegające na stwierdzeniu, gdzie

znajdowała się wskazówka w którymś miejscu w tym subiektywnym ciągu, jest samo w sobie czynnością dobrowolną i jej rozpoczęcie prawdopodobnie zabiera czas. Jest ono trudne nie tylko dlatego, że rywalizuje z innymi równoczesnymi zadaniami, ale również dlatego, że jest nienaturalne – świadomy osąd pewnego rodzaju czasowości zwykle nie odgrywa żadnej roli w regulacji zachowania, więc nie ma naturalnego znaczenia w tym ciągu. Proces interpretacji, który w końcu ustala osąd subiektywnej jednoczesności, jest sam w sobie artefaktem sytuacji eksperymentalnej i *zmienia zadanie*, a więc nie mówi nam nic istotnego o rzeczywistym umiejscowieniu w czasie zwykłych nośników reprezentacji gdziekolwiek w mózgu.

Zbyt naturalna wizja, którą musimy odrzucić, jest następująca: gdzieś głęboko w mózgu rozpoczyna się inicjacja czynności; zaczyna się jako nieświadoma intencja i powoli toruje sobie drogę do teatru, nabierając po drodze określoności i mocy, aż w końcu w chwili *t* wpada na scenę, przez którą przechodzi ciąg wizualnych reprezentacji wskazówki, przybyłych z siatkówki, coraz jaśniejszych i wyraźniej zlokalizowanych. Publiczność, czyli *ja*, dostaje zadanie stwierdzenia, która reprezentacja wskazówki była „na scenie" dokładnie w momencie, gdy kłaniała się świadoma intencja. Kiedy już zostanie zidentyfikowana, obliczony może zostać czas wyjścia wskazówki z siatkówki, jak również dystans dzielący ją od teatru oraz prędkość transmisji. W ten sposób możemy określić dokładny moment, w którym świadoma intencja wystąpiła w teatrze kartezjańskim.

Niesłychane, jak kuszący jest ten obraz. Tak łatwo go sobie wyobrazić! Wydaje się tak trafny! Czy nie jest to dokładnie to, co *musi* następować, gdy dwa procesy przebiegają jednocześnie w świadomości? Nie. W rzeczywistości *nie może* to następować, gdy jednocześnie dwa procesy przebiegają w świadomości, ponieważ nie ma takiego miejsca w mózgu. Niektórzy uważają, że niespójność *tej* wizji nie wymaga odrzucenia bezwzględnego umiejscowienia doświadczeń w czasie. Wydaje się, że istnieje konkurencyjny model początku świadomości, który unika absurdalnej wizji mózgu z centrum Kartezjusza, a jednocześnie pozwala na bezwzględne umiejscowienie w czasie. Czy świadomość mogłaby być nie kwestią przybycia do punktu, a raczej kwestią reprezentacji przekraczającej pewien próg aktywacji w całej korze lub jej znacznej części? W tym modelu element treści staje się świadomy w momencie *t* nie z powodu wkroczenia do jakiegoś funkcjonalnie zdefiniowanego i anatomicznie zlokalizowanego systemu, ale z powodu zmiany stanu w miejscu, w którym się znajduje: nabycia jakiejś własności lub zwiększenia intensywności jednej ze swoich cech ponad pewien próg.

Godna uwagi jest idea, że świadomość raczej *składa się z czynności* mózgu, a nie jest *podsystemem* mózgu (zob. np. Kinsbourne 1980; Neumann 1990; Crick i Koch 1990). Co więcej, takie zmiany trybu mogą prawdopodobnie być zmierzone w czasie przez zewnętrznych obserwatorów, zapewniając, co do zasady, unikatowy i określony szereg treści mający ten specjalny tryb. Jednak nadal jest to teatr kartezjański, jeśli twierdzimy, że prawdziwe („bezwzględne") umiejscowienie w czasie takich zmian trybu jest konstytutywne dla szeregu subiektywnego. Obraz jest trochę inny, ale następstwa są takie same. Przyznanie pewnej własności odpowiedzialności za świadomość w danym momencie to dopiero połowa sukcesu; trzeba jeszcze odróżnić tę własność odpowiedzialną za świadomość w danym momencie, a chociaż naukowcy mogą być w stanie zmierzyć to swoimi instrumentami co do milisekundy, jak ma to zrobić mózg?

My, ludzie, osądzamy równoczesność i ciągi elementów naszych przeżyć, a część z tych osądów w jakiś sposób wyrażamy, więc w którymś momencie w naszych mózgach sytuacja musi się zmienić z rzeczywistego umiejscowienia reprezentacji w czasie na reprezentacje umiejscowienia w czasie, a gdziekolwiek oraz kiedykolwiek te rozróżnienia zachodzą, od tego momentu czasowe własności reprezentacji ucieleśniających owe oceny przestają być

konstytutywne dla ich treści. Obiektywne równoczesności i ciągi elementów rozproszone na całym polu kory mózgowej nie mają żadnego funkcjonalnego znaczenia, *chyba że one również mogą być poprawnie wykryte przez mechanizmy w mózgu*. Najważniejszą kwestię możemy przedstawić w postaci pytania: Co może sprawić, że *ten* ciąg będzie strumieniem świadomości? Nie ma nikogo w środku, nikogo, kto *patrzy* na przedstawienie na ekranie rozciągniętym na całą szerokość kory, nawet jeśli takie przedstawienie jest wykrywalne przez *zewnętrznego* obserwatora. Ważne jest, w jaki sposób te treści zostają wykorzystane w procesie nieprzerwanej regulacji zachowania lub włączone do niego, a to *musi* być jedynie pośrednio ograniczone przez umiejscowienie w czasie w korze. Ważne są nie tyle czasowe własności nośników, ile czasowe własności *treści*, coś określone przez to, jak są one „wykorzystane" przez procesy następujące później w mózgu.

5. Gratka: prekognitywna karuzela Greya Waltera

Po przebrnięciu przez bardzo skomplikowane przypadki zasługujemy na poznanie czegoś dziwnego, ale *stosunkowo* łatwego do zrozumienia – czegoś, co doprowadzi do końca ten trudny rozdział. Eksperyment Libeta z umiejscawianiem w czasie przez badanych, któremu przed chwilą się przyjrzeliśmy, stworzył sztuczne i trudne zadanie osądzania, przez co wyniki nie miały znaczenia, jakiego można by się spodziewać. Ważny i wczesny eksperyment brytyjskiego neurochirurga Williama Greya Waltera (1963) nie miał tej wady. Grey Walter wykonał ten eksperyment na pacjentach, którym umieścił elektrody w korze motorycznej. Chciał sprawdzić hipotezę mówiącą, że pewne erupcje rejestrowanych czynności inicjowały działania dobrowolne. Poprosił więc pacjentów o przyglądanie się slajdom z rzutnika. Pacjent mógł dobrowolnie przyspieszyć zmianę przezrocza, wciskając odpowiedni przycisk. (Zwróćmy uwagę na podobieństwo do eksperymentu Libeta: była to „wolna" decyzja, podejmowana pod wpływem wewnętrznego poczucia nudy, ciekawości, co będzie na kolejnym slajdzie, zakłócenia uwagi lub czegokolwiek innego). Jednak wbrew temu, co sądzili pacjenci, przycisk był atrapą, w żaden sposób niepodłączoną do rzutnika! Przezrocza przesuwał tylko wzmocniony sygnał z elektrody wszczepionej do kory motorycznej pacjenta.

Można by przypuszczać, że pacjenci nie zauważyli nic dziwnego, lecz w rzeczywistości byli zszokowani efektem, ponieważ wydawało im się, jakby rzutnik przewidywał ich decyzje. Relacjonowali, że dokładnie w momencie, w którym już mieli wciskać przycisk, ale zanim tak naprawdę to postanowili, rzutnik pokazywał kolejny slajd – i wówczas wciskali przycisk, martwiąc się, że rzutnik przerzuci do przodu dwa slajdy! Według relacji Greya Waltera efekt był silny, ale najwyraźniej nigdy nie przeprowadził wymaganego kolejnego eksperymentu: wprowadzając zmienny element opóźniający, aby sprawdzić, jak długie musiałoby być opóźnienie, by wyeliminować efekt „karuzeli prekognitywnej".

Ważna różnica pomiędzy projektem Greya Waltera i Libeta jest taka, że ocena kolejności czasowej, która prowadzi do niespodzianki w eksperymencie Greya Waltera, należy do normalnego zadania monitorowania zachowania. W tym sensie jest jak ocena kolejności czasowej, dzięki której nasze mózgi rozróżniają ruch ze strony lewej na prawą od ruchu z prawej na lewą, nie jest natomiast „celowym, świadomym" osądem kolejności. W tym przypadku mózg ustawił się na „oczekiwanie" wizualnej informacji zwrotnej po udanej realizacji swojego planu przesunięcia slajdu, a ta informacja zwrotna pojawia się wcześniej, niż można by się jej spodziewać, i uruchamia alarm. Może nam to pokazać coś ważnego o rzeczywistym umiejscowieniu w czasie nośników treści oraz towarzyszących im procesach w mózgu, ale wbrew pozorom nie może nam pokazać niczego o „bezwzględnym umiejscowieniu w czasie

świadomej decyzji przesunięcia slajdu".

Załóżmy na przykład, że rozwinięcie eksperymentu Greya Waltera ujawniło, iż 300-milisekundowe opóźnienie (jak to założone przez Libeta) musi być włączone w realizację czynności, aby wyeliminować subiektywne poczucie prekognitywnej zmiany slajdu. W rzeczywistości to opóźnienie pokazałoby, że oczekiwania spowodowane decyzją o zmianie slajdu są dostrojone do tego, by uzyskać wizualną informację zwrotną 300 ms później oraz by wszcząć alarm w innych warunkach. (Jest to analogiczne do wiadomości od zszokowanego dowódcy w Kalkucie do brytyjskich władz w Londynie w następstwie bitwy pod Nowym Orleanem). Fakt, że alarm *w końcu* zostaje zinterpretowany w łańcuchu subiektywnym jako odczucie zdarzeń o zamienionej kolejności (zmiana przed wciśnięciem przycisku), nie mówi nam nic o tym, *kiedy* w rzeczywistym czasie po raz pierwszy pojawiła się świadomość decyzji o wciśnięciu przycisku. Poczucie, które zrelacjonowali badani, związane z brakiem czasu na „zawetowanie" zainicjowanego wciskania przycisku, gdy zdali sobie sprawę z tego, że „slajd już się zmienia", jest naturalną interpretacją mózgu, który (w końcu) bierze pod uwagę wszystkie treści dostępne w różnych momentach i włącza je do narracji. Czy to poczucie pojawiło się w pierwszym momencie świadomości intencji (wówczas efekt wymaga długiego opóźnienia do „czasu pokazu", jak w interpretacji stalinowskiej), czy była to retrospektywna reinterpretacja w innym razie dezorientującego *faktu dokonanego* (w interpretacji orwellowskiej)? Mam nadzieję, że teraz już jest jasne, iż założenia tego pytania po prostu je dyskwalifikują.

6. Niedopowiedzenia

Być może nadal chcesz zaprzeczyć temu, jakoby argumenty przedstawione w tym rozdziale nie były wystarczająco silne, aby obalić oczywistą prawdę, że nasze przeżycia dotyczące zdarzeń następują w tej samej kolejności, w jakiej ich doświadczamy. Jeśli ktoś myśli „jeden, dwa, trzy, cztery, pięć", jego myślenie o „jeden" następuje przed pomyśleniem o „dwa" i tak dalej. Ten przykład ilustruje na ogół prawdziwą tezę, a wyjątki od niej wydają się niemożliwe, dopóki ograniczamy naszą uwagę do psychologicznych zjawisk o „zwykłym", makroskopijnym trwaniu. Jednak eksperymenty, którym się przyglądaliśmy, zajęły się zdarzeniami ograniczonymi niezwykle wąskimi ramami czasowymi trwającymi kilkaset milisekund. Na tym poziomie standardowe założenia przestają być użyteczne.

Każde zdarzenie w mózgu ma określoną lokalizację czasoprzestrzenną, ale pytanie: „kiedy dokładnie bodziec stał się świadomy?" zakłada, że jedno z tych zdarzeń jest uświadomieniem sobie bodźca bądź sprowadza się do niego. To tak, jakby zapytać: „Kiedy dokładnie Imperium Brytyjskie zostało poinformowane o rozejmie w wojnie roku 1812?". Gdzieś pomiędzy 24 grudnia 1814 a środkiem stycznia 1815 roku – to jest pewne, jednak tak naprawdę nie jesteśmy w stanie określić dokładnego dnia i godziny. Nawet jeśli udałoby nam się podać dokładny czas, w którym różne władze Imperium zostały o tym fakcie poinformowane, żaden z tych momentów nie może być uznany za czas, w którym się ono o tym dowiedziało. Podpisanie rozejmu było pewnym oficjalnym, intencjonalnym aktem Imperium, lecz udział sił brytyjskich w wojnie pod Nowym Orleanem był kolejnym takim aktem, wykonanym z założeniem, że do żadnego rozejmu nie doszło. Można by słusznie zauważyć, że z zasady dotarcie informacji do Whitehall lub pałacu Buckingham w Londynie powinno być uważane za oficjalny moment, w którym Imperium o czymś się dowiaduje, ponieważ jest to jego „centrum nerwowe". Kartezjusz uważał, że to szyszynka była takim centrum nerwowym w mózgu, ale się mylił. Postrzeganie i kontrola – a stąd świadomość – są rozmieszczone w różnych częściach mózgu, zatem żaden moment nie może zostać uznany za ten, w którym następowało świadome zdarzenie.

W tym rozdziale próbowałem wykorzenić złe nawyki myślowe, odrywając je od ich wyimaginowanych „fundamentów", i zastąpić je lepszymi sposobami myślenia, jednak po drodze musiałem zostawić wiele niedopowiedzeń. Podejrzewam, że najbardziej kuszące jest metaforyczne twierdzenie, że „badanie" jest czymś, co „wywołuje narracje". Uważam, że umiejscowienie w czasie pytań kontrolnych przez badaczy może mieć znaczący wpływ na systemy reprezentacji używane przez mózg. Ale pośród tych, którzy mogą skierować takie pytania kontrolne do osoby badanej, jest sama ta osoba. Jeśli interesuje cię, kiedy coś sobie uświadamiasz, twoje własne badania i dociekliwość ustalą granice nowych okien kontrolnych i w ten sposób przesuną granice procesów w nich przebiegających.

Rezultatami badań zewnętrznych są zwykle jakiegoś rodzaju akty mowy, a te *wyrażają sądy* na temat różnorodnych treści świadomości, będące przedmiotem późniejszych interpretacji. Rezultaty badań na samym sobie to elementy *tej samej kategorii semantycznej* – nie „przedstawienia" (w teatrze kartezjańskim), lecz *sądy* dotyczące tego, co osobie badanej się wydaje, oceny, które następnie ta osoba może sama zinterpretować, zadziałać zgodnie z nimi, zapamiętać. W obu przypadkach owe zdarzenia ustalają interpretacje tego, co doświadczyła ta osoba, a zatem stanowią stałe punkty w ciągu subiektywnym. Jednak w modelu wielokrotnych szkiców pozbawione sensu jest pytanie, *czy oprócz* takiego sądu oraz wcześniejszych rozróżnień, na których jest on oparty, odbyło się przedstawienie materiałów do interpretacji, aby mógł je skontrolować główny sędzia, publiczność w teatrze kartezjańskim. To nadal trudno zrozumieć, a co dopiero zaakceptować. Musimy pokazać, że do tego wniosku wiodą jeszcze inne drogi.

Rozdział 7

Ewolucja świadomości

Wszystko jest takie, jakie jest, ponieważ takie się stało.
D'Arcy Thompson, 1917

1. Wewnątrz czarnej skrzynki świadomości

Teoria zarysowana w poprzednim rozdziale częściowo wskazuje, jak świadomość może się pojawiać w ludzkich mózgach, jednak jej zasadniczy wydźwięk był negatywny: obalenie dyktatorskiej idei teatru kartezjańskiego. *Zaczęliśmy* ją zastępować modelem pozytywnym, ale nie zaszliśmy za daleko. Aby dalej go rozwijać, musimy zmienić dziedzinę i podejść do złożoności świadomości od innej strony: od ewolucji. Ludzka świadomość nie istnieje od zawsze, więc musiała wyłonić się z innych zjawisk, które świadomością nie były. Być może, gdy spojrzymy na to, co musiało – lub mogło – warunkować to przejście, lepiej zrozumiemy złożoność i jej rolę w powstawaniu tego końcowego zjawiska.

Neuronaukowiec Valentino Braitenberg w swojej eleganckiej książeczce *Vehicles: Essays in Synthetic Psychology* (1984) opisuje szereg coraz bardziej skomplikowanych mechanizmów autonomicznych, stopniowo powstających ze śmiesznie prostych i kompletnie bezdusznych przyrządów oraz stających się (wyobrażonymi) jednostkami, które imponują realnością biologiczną i psychologiczną. To ćwiczenie wyobraźni działa ze względu na to, co Braitenberg nazywa „prawem analizy wstępującej i syntezy zstępującej": o wiele łatwiej wyobrazić sobie zachowanie (i jego następstwa) urządzenia syntetyzowanego poniekąd „od wewnątrz", niż próbować analizować zewnętrzne zachowanie „czarnej skrzynki" i domyślać się, co musi się dziać w środku.

Dotychczas traktowaliśmy świadomość jako coś na kształt czarnej skrzynki. Wyciągnęliśmy jej „zachowanie" (= fenomenologię) jako coś oczywistego i zastanawialiśmy się, jaki rodzaj ukrytych mechanizmów w mózgu mógłby to zachowanie wyjaśnić. Teraz odwróćmy tę strategię: pomyślmy o ewolucji mechanizmów mózgowych odpowiedzialnych za różne kwestie i zobaczmy, czy to, co się wyłania, może być wiarygodnym mechanizmem, który wyjaśniałby dziwne „zachowania" naszych świadomych mózgów.

Istnieje wiele teorii – czy raczej spekulacji – mówiących o ewolucji ludzkiej świadomości, a pierwsza z nich to rozważania Darwina w *O pochodzeniu człowieka* (1871/1959). W przeciwieństwie do innych wyjaśnień naukowych, wyjaśnienia ewolucyjne są przede wszystkim narracjami rozpoczynającymi się w czasie, w którym coś nie istniało, a kończącymi się w momencie zaistnienia danego zjawiska, prowadzącymi nas przez serię kroków wyjaśnianych po drodze. Zamiast przedstawiać wszystkie dotychczas wymyślone narracje tego typu, systematycznie i naukowo je omawiając, opowiem jedną historię, zapożyczając swobodnie różne elementy od innych teoretyków i koncentrując się na kilku pominiętych punktach, które pomogą nam pokonać przeszkody w zrozumieniu, czym jest świadomość. Aby dobrze opowiedzieć historię i aby była ona stosunkowo krótka, oparłem się pokusie włączenia dosłownie kilkudziesięciu fascynujących wątków pobocznych i trzymałem na wodzy standardowy instynkt filozofa, nakazujący prezentowanie wszystkich argumentów za i przeciw elementom, które

włączam, oraz tym, które odrzucam. Przyznaję, że rezultat przypomina trochę streszczenie *Wojny i pokoju* w stu słowach, ale tylko na tyle wystarczy nam miejsca[51].

Historia, którą musimy opowiedzieć, jest analogiczna do innych historii opowiadanych od niedawna w biologii. Można ją porównać na przykład do historii o pochodzeniu płci. Istnieje obecnie wiele organizmów bezpłciowych, które tak też się rozmnażają, ale był czas, kiedy to żadne organizmy nie miały płci męskiej i żeńskiej. W jakiś sposób, poprzez wyobrażalną serię kroków, niektóre z tych organizmów musiały przekształcić się w takie, które mają płeć, aż w końcu ewolucja stworzyła nas. Jakie warunki były sprzyjające czy konieczne do zaistnienia tych innowacji? Jest to jeden z najgłębszych problemów współczesnej teorii ewolucji[52].

Istnieje elegancka zbieżność między tymi dwoma zagadnieniami, pochodzeniem płci i pochodzeniem świadomości. Nie ma właściwie nic *seksownego* (w znaczeniu ludzkim) w życiu seksualnym ptaków, ostryg i innych prostych form życia, lecz w ich mechanicznej i pozornie pozbawionej radości rutynie możemy rozpoznać fundamenty i zasady naszego o wiele bardziej podniecającego świata seksu. W podobny sposób nie ma nic szczególnie *jaźniowego* (jeśli mogę stworzyć taki termin) w prymitywnych prekursorach świadomych jednostek ludzkich, jednak są one podstawą naszych typowo ludzkich innowacji i naszej złożoności. Konstrukcja naszych świadomych umysłów jest rezultatem trzech kolejnych procesów ewolucyjnych, nakładających się na siebie, a każdy jest wyraźnie szybszy i potężniejszy od swojego poprzednika, ale żeby zrozumieć tę piramidę procesów, musimy zacząć od początku.

2. Początki

Scena pierwsza: Narodziny granic i racji

Na początku nie było racji; były tylko przyczyny. Nic nie miało celu, nic nie miało więcej niż funkcję; nie było na świecie żadnej teleologii. Wyjaśnienie jest proste: nie istniało nic, co miałoby swoje interesy. Jednak po upływie tysiącleci pojawiły się zwykłe *replikatory* (Dawkins 1976/1996; zob. Monod 1972/1979, rozdz. 1). *One* nie miały pojęcia o swoich interesach, a być może właściwiej byłoby powiedzieć, że wcale ich nie miały, jednak my, spoglądając z naszego niemal boskiego punktu widzenia na ich początki, możemy nieprzypadkowo przypisać im pewnego rodzaju zainteresowania – będące pochodnymi podstawowego „interesu", jakim była autoreplikacja. Innymi słowy, być może nie miało to znaczenia, było kwestią nieistotną, nie było dla nikogo i niczego ważne, czy osiągną one sukces replikacyjny, czy nie (choć wydaje się, iż możemy być wdzięczni, że im się udało), ale przynajmniej możemy im przypisać interesy mogące zaistnieć pod pewnym warunkiem. *Jeśli* te proste replikatory mają przeżyć i się mnożyć, a więc przetrwać w obliczu narastającej entropii, ich środowisko musiało spełniać pewne warunki: warunki sprzyjające replikacji muszą być obecne, a przynajmniej częste.

Ujmując sprawę bardziej antropomorficznie, jeśli te proste replikatory chcą nadal się mnożyć, powinny mieć nadzieję i walczyć o różne rzeczy; powinny unikać rzeczy „złych" i poszukiwać „dobrych". Kiedy na scenę wkracza jednostka zdolna do zachowania odsuwającego w czasie, choćby tylko w sposób najbardziej prymitywny, poprzez rozpad i rozkład, wraz z tą zdolnością sprowadza na świat swoje „dobro". Innymi słowy, stwarza punkt widzenia, z którego wydarzenia mogą być z grubsza podzielone na korzystne, niekorzystne i neutralne. Natomiast jej własne, wrodzone skłonności, aby poszukiwać pierwszych, unikać drugich i ignorować trzecie, zasadniczo definiują te trzy klasy. Gdy tak oto istota zaczyna mieć swoje interesy, a świat i zachodzące w nim zdarzenia zaczynają stanowić dla niej *racje* – bez względu na to, czy istota jest w stanie całkowicie je rozpoznać (Dennett 1984a). Pierwsze racje istniały, zanim zostały

rozpoznane. Tak naprawdę pierwszym problemem, któremu stawiła czoło pierwsza istota stawiająca czoło problemom, było nauczenie się, jak rozpoznawać i działać na podstawie racji, które wytworzyło jej własne istnienie.

Gdy tylko coś zaczyna interesować się zachowaniem samego siebie, ważne stają się granice, ponieważ gdy zamierzasz się chronić, nie chcesz trwonić wysiłków, próbując ocalić cały świat: wytyczasz granice. Jednym słowem, stajesz się *egoistą*. Ten pierwotny rodzaj egoizmu (który, jako forma pierwotna, nie ma większości odcieni naszej formy egoizmu) jest jedną z oznak życia. To, gdzie kończy się jeden kawałek granitu, a zaczyna drugi, pozostaje bez znaczenia; granica jego pęknięcia może być rzeczywista, ale nie ma nic, co ochroniłoby terytorium, przesunęło tę granicę lub wycofało się z niej. „Ja przeciwko światu" – to rozróżnienie wszystkiego tego, co znajduje się wewnątrz zamkniętej granicy, oraz tego, co jest w świecie zewnętrznym – leży u sedna wszystkich procesów biologicznych, nie tylko połykania i wydalania, oddychania i transpiracji. Weźmy na przykład układ immunologiczny z milionami różnych przeciwciał ustawionych w obronie ciała przeciwko milionom różnorodnych zewnętrznych intruzów. Ta armia musi rozwiązać fundamentalny problem rozpoznawczy: odróżnienia samego siebie (i swoich przyjaciół) od wszystkiego innego. Problem ten został rozwiązany z grubsza w ten sam sposób, z którego korzystają państwa i ich armie: przez standaryzowane, zmechanizowane procedury identyfikacyjne – miniaturowe paszporty i celnicy to kształty cząsteczek oraz detektory tych kształtów. Należy zaznaczyć, że ta armia przeciwciał nie ma generałów, głównej centrali dowodzącej z planem walki ani nawet opisu wroga: przeciwciała reprezentują swoich wrogów tylko w sposób, w jaki milion zamków reprezentuje klucze, które je otwierają.

Musimy zwrócić uwagę na jeszcze kilka faktów ujawniających się już na tym wczesnym etapie. Ewolucja zależy od historii, ale Matka Natura nie jest snobką i pochodzenie nie robi na niej wrażenia. Nie ma znaczenia, gdzie czy jak dany organizm zdobył swoje siły; nie szata zdobi człowieka. Oczywiście, o ile dzisiaj wiemy, pochodzenie wczesnych replikatorów było w każdym przypadku takie samo: każdy z nich był wynikiem takiej czy innej ślepej, przypadkowej selekcji. Gdyby jednak podróżujący w czasie superinżynier umieścił replikatora-robota w tym środowisku i gdyby jego zdolności były takie same lub lepsze niż zdolności jego naturalnej konkurencji, jego potomkowie mogliby teraz być wśród nas – mogliby nawet *być* nami! (Dennett 1987a, 1990b).

Dobór naturalny nie może powiedzieć, dlaczego jakiś system stał się właśnie taki, ale nie znaczy to, że nie mogą istnieć głębokie różnice między systemami „zaprojektowanymi" przez dobór naturalny a tymi zaprojektowanymi przez inteligentnych inżynierów (Langton, Hogeweg, w: Langton 1989). Na przykład projektanci, cechując się dalekowzrocznością, ale i klapkami na oczach, zwykle stwierdzają, że ich projektom zagrażają nieprzewidziane efekty uboczne i oddziaływania, więc starają się przed nimi chronić, dając każdemu z elementów w systemie jedną funkcję i izolując go od innych elementów. Inaczej jest z Matką Naturą (procesem doboru naturalnego), która jest znana z krótkowzroczności i braku celów. Nie przewiduje w ogóle, więc nie może się martwić o nieprzewidziane efekty uboczne. Nie „próbując" ich uniknąć, wypróbowuje projekty, z których wiele cierpi na efekty uboczne; większość tych projektów jest beznadziejna (zapytaj jakiegokolwiek inżyniera), jednak od czasu do czasu pojawia się *przypadkowy, ale korzystny efekt uboczny*: co najmniej dwa niepowiązane ze sobą systemy funkcyjne współdziałają i dają premię: wiele funkcji jednego elementu. Wielofunkcyjność nie jest nieznana w przedmiotach stworzonych przez ludzi, lecz są one rzadkie; w naturze są wszędzie i – jak zobaczymy – jednym z powodów, dla których teoretycy mają taki problem z odkryciem wiarygodnych projektów świadomości w mózgu, jest to, że zwykle uważają, iż

elementy mózgu pełnią tylko jedną funkcję[53].

Znamy już podstawy. Możemy teraz wyjaśnić następujące fakty pierwotne:

(1) Istnieją racje, które można rozpoznawać.

(2) Jeśli są racje, istnieją też punkty widzenia, z których perspektywy można te racje rozpoznać i ocenić.

(3) Każdy podmiot działający musi odróżniać „tu, wewnątrz" od „świata zewnętrznego".

(4) Każde rozpoznanie musi koniec końców być zrealizowane przez wiele „ślepych, mechanicznych" procedur.

(5) Wewnątrz określonej granicy nie musi zawsze istnieć dyrektor czy centrala.

(6) W naturze nie szata zdobi człowieka; pochodzenie nie ma znaczenia.

(7) W naturze elementy często pełnią wiele funkcji w ramach jednego organizmu.

Widzieliśmy już konsekwencje tych pierwotnych faktów w poszukiwaniach ostatecznego „punktu widzenia świadomego obserwatora" i w kilku przykładach, w których homunkulusa zamieniliśmy na proste mechanizmy czy ich zespoły. Jednak, jak mogliśmy zaobserwować, punkt widzenia świadomego obserwatora nie jest tożsamy z pierwotnymi punktami widzenia pierwszych replikatorów, które dzieliły swoje światy na to, co dobre, i to, co złe, ale jest ich bardziej wysublimowanym potomkiem. (W końcu nawet rośliny mają swój punkt widzenia w tym pierwotnym sensie).

Scena druga:
Nowe i lepsze sposoby tworzenia przyszłości

Jedną z najgłębszych, najogólniejszych funkcji organizmów żywych jest patrzeć przed siebie, tworzyć przyszłość, jak określił to Paul Valéry.

François Jacob, 1982, s. 66

Przewidzieć przyszłość pewnej krzywej to znaczy wykonać pewne operacje na jej przeszłości. Idealnego operatora przewidywania nie da się skonstruować w postaci aparatu, jednakże istnieją pewne operatory, zbliżone nieco do idealnych i dające się urzeczywistnić w postaci urządzeń, których budowa leży w zasięgu naszych możliwości.

Norbert Wiener, 1948/1971, s. 28–29

W poprzednim rozdziale mimochodem wspomniałem, że głównym zadaniem mózgów jest wytwarzanie przyszłości i twierdzenie to zasługuje na trochę więcej uwagi. Aby przeżyć, organizm musi się uzbroić (jak drzewo czy małż) i być dobrej myśli lub opracować metody unikania szkód i przemieszczania się w bezpieczniejsze miejsca. Kto zdecyduje się na to drugie, ten musi zmierzyć się z pierwotnym problemem, który nieustannie rozwiązywać musi każdy *podmiot działający*:

Co mam teraz robić?

Do rozwiązania tego problemu potrzebny jest układ nerwowy umożliwiający ci *regulowanie* działań w czasie i przestrzeni. Młoda żachwa przemierza morze w poszukiwaniu

odpowiedniej skały lub kawałka rafy, aby się w nią wczepić i założyć dom na całe życie. Do wypełnienia tego zadania ma elementarny układ nerwowy. Gdy znajdzie miejsce i się zakorzeni, nie potrzebuje już swojego mózgu, więc go zjada! (Co przypomina habilitację)[54]. Kluczem regulacji jest umiejętność *śledzenia*, a nawet *przewidywania* ważnych właściwości środowiska, zatem wszystkie mózgi to w istocie *maszyny przewidujące*. Muszla małża to świetna zbroja, ale nie zawsze może być zamknięta; wbudowany odruch, który nagle zamyka muszlę, jest prostacki, lecz skuteczny, gdyż przewiduje krzywdę lub jej unika.

Jeszcze bardziej prostackie są reakcje wycofania i zbliżania u najprostszych organizmów, a łączą się ze źródłami dobra i zła najbardziej bezpośrednio: *dotykiem*. W zależności od tego, czy dotknięta rzecz jest dla nich zła, czy dobra, wycofują się lub ją pochłaniają (w samą porę, jeśli mają szczęście). Dzieje się to, gdyż organizmy te są skonstruowane tak, że kontakt z dobrą lub złą cechą wywołuje odpowiedni ruch. Jak zobaczymy, fakt ten jest podstawą niektórych najstraszniejszych i najsmaczniejszych (dosłownie) cech świadomości. Na początku *wszystkie* „sygnały" spowodowane przez środowisko znaczyły „uciekaj!" lub „do dzieła!" (Humphrey 1992).

Na tym wczesnym etapie żaden układ nerwowy nie miał możliwości wykorzystania chłodniejszej bądź obiektywnej „wiadomości", która by go zaledwie *informowała* neutralnie o pewnym stanie. Jednak takie proste układy nerwowe nie są w stanie znaleźć punktu oparcia w świecie. Są on jedynie zdolne do czegoś, co moglibyśmy nazwać *„antycypacją proksymalną"*: do zachowania, które jest odpowiednie do tego, co wydarzy się w *najbliższej* przyszłości. Lepsze mózgi potrafią wydobyć więcej informacji szybciej i użyć ich przede wszystkim do *uniknięcia* szkodliwego kontaktu lub do *poszukiwania* pokarmu (oraz możliwości łączenia się w pary, w momencie gdy pojawiła się płeć).

Stawiając czoło zadaniu zbudowania użytecznej przyszłości z osobistej przeszłości, my, organizmy, staramy się dostać coś za darmo (a przynajmniej w dobrej cenie): odnaleźć prawa rządzące światem – a jeśli takowych nie ma, to znaleźć przybliżone prawa rządzące światem – cokolwiek, co da nam przewagę. Z pewnej perspektywy wydaje się zupełnie niesłychane, że my, organizmy, w ogóle znajdujemy punkt oparcia w naturze. Czy istnieje jakaś głęboka racja, dla której natura powinna nam pomagać lub ujawniać swoje prawa na drodze zwykłej obserwacji? Każdy przydatny generator przyszłości jest czymś w rodzaju sztuczki – doraźnym układem, który akurat w większości przypadków działa, szczęśliwym trafieniem jakiejś prawidłowości w świecie, którą można śledzić. Każdemu takiemu organizmowi, który ma szczęście w przewidywaniu, przysługuje oczywiście nagroda ze strony Matki Natury, jeśli tylko pogłębi w ten sposób swoją przewagę.

Na krańcu minimalistycznym mamy zatem stworzenia reprezentujące jak najmniej: jedynie tyle, by *świat mógł je czasem ostrzec*, kiedy zaczynają coś robić źle. Stworzenia, które kierują się tą strategią, nie planują. Idą przed siebie, a gdy coś zacznie je boleć, „wiedzą", że muszą się wycofać, ale jedynie tyle potrafią.

Kolejny etap to krótkoterminowa antycypacja – na przykład umiejętność uchylania się przed lecącą cegłą. Tego rodzaju talent do antycypowania jest zazwyczaj „wbudowany" – należy do wrodzonej maszynerii projektowanej przez całe wieki, a śledzącej swego rodzaju wyjątkową regularność, którą można dostrzec między *rzeczami zbliżającymi się* a *rzeczami, które nas uderzają*. Uchylenie się przed zbliżającym się przedmiotem jest głęboko zakorzenione na przykład w ludziach i można je zaobserwować u noworodków (Yonas 1981), co jest darem od naszych przodków, których zmarli kuzyni nie potrafili się wystarczająco uchylić. Czy sygnał „coś się zbliża" *znaczy* „uchyl się!"? Cóż, takie jest jego protoznaczenie; jest on bezpośrednio połączony z mechanizmem uchylania się.

Dostaliśmy też inne dary. Nasze układy wzrokowe, tak jak u wielu innych zwierząt, nawet ryb, są wrażliwe na wzorce mające pionową oś symetrii. Braitenberg twierdzi, że to prawdopodobnie dlatego, iż w naturalnym środowisku naszych odległych przodków (na długo zanim powstały fasady kościołów i wiszące mosty) niemalże jedynymi rzeczami, które miały taką pionową oś symetrii, były inne zwierzęta, i to tylko patrzące się na ciebie. Zatem nasi przodkowie byli wyposażeni w najcenniejszy system alarmowy, który był uruchamiany (głównie), gdy inne zwierzę *patrzyło* na nich (Braitenberg 1984)[55]. Identyfikacja drapieżnika w (przestrzennej) odległości (w odróżnieniu od czekania, aż poczujesz jego zęby wbijające się w ciebie) również jest rodzajem przewidywania dystalnego *czasowo*: daje ci przewagę w unikaniu [niebezpieczeństw].

Ważnym faktem dotyczącym takich mechanizmów jest to, że ich rozróżnienia są bardzo zgrubne; idą na kompromis między czymś, co można by nazwać *prawdą i trafnością ujęcia*, a *szybkością i oszczędnością*. Niektóre z bodźców, które wyzwalają detektor symetrii, nie mają właściwie żadnego znaczenia dla organizmu: rzadkie symetryczne drzewo czy krzak lub (współcześnie) wiele przedmiotów wyprodukowanych przez ludzi. Zatem klasa rzeczy rozróżnianych przez ten mechanizm to po prostu kolorowa zbieranina – zdominowana przez zwierzęta patrzące się w moim kierunku, ale dopuszczająca nieograniczenie wielką liczbę fałszywych alarmów (w odniesieniu do *tego* komunikatu). Poza tym nie wszystkie oraz nie tylko wzorce o pionowej symetrii uruchomią ten proces; niektóre wzorce pionowe z takiego czy innego powodu tego nie zrobią, a poza tym wystąpią fałszywe alarmy; jest to cena, którą należy zapłacić za szybki, tani, przenośny mechanizm, cena, którą chętnie płacą organizmy w swoim narcyzmie (Akins 1989). To fakt oczywisty, jednak niektóre z jego następstw dla świadomości nie są z początku tak oczywiste. (W rozdziale 12 stanie się to istotne, gdy będziemy zadawać pytania typu: Jakie właściwości wykrywamy za pomocą naszego wzroku, rozróżniającego kolory? Co mają ze sobą wspólnego rzeczy czerwone? A nawet: dlaczego świat wydaje nam się taki, a nie inny?)

Dowiedzenie się (choćby i omylne) o tym, że inne zwierzę patrzy na ciebie, jest w świecie przyrody niemal zawsze zdarzeniem istotnym. Jeśli zwierzę to nie chce cię zjeść, to może być potencjalnym partnerem seksualnym bądź rywalem w zdobyciu takiego partnera, czy też ofiarą, która zdała sobie sprawę z tego, że się zbliżasz. Ten alarm powinien następnie włączyć analizę „przyjaciel, wróg czy pokarm?", aby organizm mógł rozróżnić komunikaty typu: „przedstawiciel tego samego gatunku patrzy na ciebie!", „zostałeś namierzony przez drapieżnika!" i „twoja kolacja zaraz ucieknie!". U niektórych gatunków (na przykład u pewnych ryb) detektor symetrii pionowej jest skonstruowany tak, że powoduje nagłe przerwanie wykonywanej czynności, co jest znane jako *reakcja orientacyjna*.

Psycholog Odmar Neumann (1990) uważa, że reakcje orientacyjne są biologicznymi odpowiednikami alarmu na statku „wszystkie ręce na pokład!". Większość zwierząt, tak jak my, wykonuje niektóre czynności rutynowo, „na autopilocie", wykorzystując mniej niż pełną wydolność, a czynności te są w istocie pod kontrolą wyspecjalizowanych podsystemów znajdujących się w ich mózgach. Kiedy wszczęty jest wyspecjalizowany alarm (jak na przykład alarm zbliżającej się istoty czy alarm symetrii pionowej w człowieku) lub alarm ogólny z powodu czegoś nagłego i zaskakującego (czy po prostu niespodziewanego), układ nerwowy zwierzęcia zostaje zmobilizowany, aby radzić sobie z możliwością nagłego wypadku. Zwierzę przerywa poprzednie czynności i szybko rozgląda się czy też odświeża informacje, co daje każdemu organowi zmysłowemu możliwość wniesienia czegoś do puli dostępnych i rzetelnych informacji. Zostaje stworzona *tymczasowo* scentralizowana arena kontroli dzięki zwiększonej aktywności neuronalnej – wszystkie kanały są otwarte przez krótką chwilę. Jeśli wynikiem

zebranych informacji jest wszczęcie „drugiego alarmu", całe ciało zwierzęcia zostaje zmobilizowane przez nagły przypływ adrenaliny. Jeśli nie, zwiększona aktywność wkrótce się cofa, załoga po służbie wraca do łóżka, a specjaliści powracają do swoich funkcji kontrolnych. Te krótkie epizody przerywania i zwiększonej czujności nie są same w sobie przypadkami „przytomnej świadomości" (jak czasem mówią niektórzy) typu ludzkiego, a przynajmniej nie są nimi koniecznie, są jednak prawdopodobnie koniecznymi ewolucyjnymi prekursorami świadomych stanów.

Neumann twierdzi, że reakcje orientacyjne powstały jako reakcje na sygnały alarmowe, ale okazały się tak przydatne, wywołując ogólne odświeżanie, że zwierzęta zaczęły wchodzić w ten tryb orientacyjny coraz częściej. Ich układ nerwowy potrzebował trybu „wszystkie ręce na pokład!", a gdy już został on osiągnięty, częstsze uruchamianie go kosztowało niewiele lub nic, opłaciło się zaś z nawiązką, zapewniając lepsze informacje o stanie środowiska czy o stanie samego zwierzęcia. Można by powiedzieć, że stało się to nawykiem, już nie tylko pod kontrolą zewnętrznych bodźców, ale inicjowanym wewnątrz (podobnie jak regularne ćwiczenia przeciwpożarowe).

Regularna czujność stopniowo zmieniła się w regularną *eksplorację* i zaczęła ewoluować nowa strategia behawioralna: strategia pozyskiwania informacji „dla samej informacji", tylko na wypadek, gdyby mogła się okazać któregoś dnia cenna. Strategia ta przyciągnęła większość ssaków, szczególnie naczelnych, które rozwinęły bardzo ruchome oczy, a te dzięki ruchom sakkadowym zapewniały niemal nieprzerwane przeszukiwanie świata. Spowodowało to fundamentalną zmianę w ekonomii organizmów, które zrobiły ten krok: narodziny ciekawości, czyli głodu epistemologicznego. Zamiast gromadzić informacje jedynie na daną chwilę, do wykorzystania natychmiast, zaczęły stawać się czymś, co psycholog George Miller nazwał *„informacjofagami"*: organizmami spragnionymi nowych informacji o świecie, który zamieszkują (oraz o sobie). Jednak nie stworzyły i nie zastosowały całkowicie nowych systemów gromadzenia informacji. Jak to zwykle bywa w ewolucji, zmontowały te nowe systemy ze sprzętu zapewnionego im przez ich przodków. Ta historia zostawiła po sobie ślady, zwłaszcza na emocjonalnych czy afektywnych zabarwieniach świadomości, gdyż mimo że bardziej zaawansowane stworzenia stały się „niezainteresowanymi" zbieraczami informacji, ich „mechanizmy raportowania" były tylko inaczej zastosowanymi ośrodkami ostrzegającymi i dopingującymi ich przodków, nigdy nie wysyłając żadnej informacji „wprost", ale zawsze dodając pewien szczątkowy pozytywny lub negatywny „wydźwięk" redaktorski do każdej dostarczanej przez siebie informacji. Usuwając cudzysłów i metafory: wrodzone połączenia stanów informujących z wycofaniem oraz pochłanianiem, unikaniem i wzmacnianiem nie zostały zniszczone, a jedynie osłabione i gdzie indziej ukierunkowane. (Wrócimy do tej kwestii w rozdziale 12).

U ssaków temu ewolucyjnemu wzrostowi sprzyjał podział pracy w mózgu, który wytworzył dwa wyspecjalizowane obszary: (z grubsza) *grzbietowy* i *brzuszny*. (Hipoteza, którą przedstawiam dalej, została postawiona przez neuropsychologa Marcela Kinsbourne'a). Część grzbietowa mózgu jest odpowiedzialna za „bieżące" *pilotowanie*, mające utrzymać statek (ciało organizmu) z dala od niebezpieczeństwa; tak jak procedury „wykrywania zderzeń" wbudowane w gry wideo, musiało ono niemalże bez przerwy monitorować środowisko w poszukiwaniu zbliżających się i oddalających przedmiotów, a ogólnie rzecz biorąc, było odpowiedzialne za unikanie wpadania organizmu na przedmioty i spadania z urwisk. Pozwoliło to brzusznej części mózgu skoncentrować się na identyfikacji różnorodnych przedmiotów w świecie; mogła ona pozwolić sobie na szczegółową obserwację konkretów i ich względnie powolną i szeregową analizę, ponieważ mogła opierać się na systemie brzusznym, który chronił ciało przed

niebezpieczeństwami. U naczelnych, według spekulacji Kinsbourne'a, podział grzbietowo-brzuszny w drodze dalszej ewolucji przekształcił się w znane specjalizacje prawej i lewej półkuli: globalną, czasoprzestrzenną prawą półkulę i bardziej skoncentrowaną, analityczną, szeregową półkulę lewą.

Śledzimy zaledwie jeden wątek w ewolucyjnej historii układu nerwowego i skorzystaliśmy z najbardziej podstawowego mechanizmu ewolucyjnego: doboru konkretnych *genotypów* (kombinacji genów), które okazały się odpowiedzialne za lepiej przystosowane jednostki (*fenotypy*) niż inne genotypy. Organizmy, które mają na tyle szczęścia, że rodzą się lepiej przygotowane, zwykle wytwarzają więcej mogących przeżyć potomków, a zatem lepsze cechy wrodzone (*hard-wired*) rozprzestrzeniają się w populacji. Naszkicowaliśmy rozwój w przestrzeni konstrukcji od najprostszych wyobrażalnych detektorów dobra i zła do zorganizowanych architektur złożonych z takich mechanizmów, mających w rezultacie zdolność tworzenia przydatnych antycypacji we względnie stabilnych i przewidywalnych środowiskach.

W trzeciej fazie naszego opowiadania musimy wprowadzić istotną innowację: pojawienie się indywidualnych fenotypów, których wnętrze nie jest całkowicie wrodzone, a raczej zmienne i *plastyczne*, przez co mogą uczyć się one w czasie swojego życia. Pojawienie się plastyczności w układach nerwowych nastąpiło (z grubsza) w tym samym czasie, co ulepszenia, o których właśnie mówiliśmy, i wytworzyło dwa nowe ośrodki, w których mogła zadziałać ewolucja, ze znacznie większą szybkością niż sama ewolucja genetyczna na drodze mutacji genów i doboru naturalnego. Niektóre złożone cechy ludzkiej świadomości są rezultatem tych nadal pojawiających się ulepszeń w owych ośrodkach, więc potrzebujemy jasnej, choć może i elementarnej, wizji związków między nimi a leżącym u ich podstaw procesem ewolucji genetycznej.

3. Ewolucja w mózgach a efekt Baldwina

Wszyscy zakładamy, że przyszłość będzie podobna do przeszłości – jest to podstawowa, choć niedowodliwa zasada każdego naszego wniosku indukcyjnego, jak zauważył David Hume. Matka Natura (budowniczy konstrukcji realizowanych w ramach doboru naturalnego) zakłada to samo. W wielu aspektach warunki pozostają niezmienione: grawitacja nadal oddziałuje, woda nadal paruje, organizmy muszą uzupełniać i chronić wodę w swoim ciele, zbliżające się przedmioty nadal zajmują coraz większe fragmenty siatkówki i tak dalej. W przypadku tak ogólnych zasad Matka Natura daje rozwiązania długoterminowe: wrodzone detektory grawitacyjne wskazujące kierunek w górę, wrodzone alarmy pragnienia, wrodzone obwody uchylania się przed zbliżającymi się przedmiotami. Inne warunki są zmienne, ale w sposób przewidywalny, cyklicznie, a Matka Natura odpowiada na nie, tworząc inne wrodzone urządzenia, na przykład mechanizmy zimowego przyrostu futra uruchamiane przez zmianę temperatury czy wrodzone budziki, które zarządzają cyklami budzenia się i zasypiania zwierząt nocnych i dziennych. Jednak czasem Matka Natura ani w ogóle nikt raczej nie może przewidzieć korzystnych i szkodliwych warunków otoczenia – są one *chaotycznymi* procesami lub zależą od takich procesów (Dennett 1984a, s. 109 i nast.). W takich przypadkach żaden stereotypowy projekt nie przystosuje się na każdą możliwą okoliczność, więc na lepszej pozycji będą te organizmy, które potrafią się w pewnym stopniu *przekonstruować*, aby stawić czoło napotkanym warunkom. Czasem takie przekonstruowanie nazywane jest „uczeniem się", a czasem po prostu *rozwojem*. Linia oddzielająca je od siebie jest sporna. Czy ptaki *uczą się* latać? Czy uczą się śpiewać? (Nottebohm 1984; Marler i Sherman 1983). A czy uczą się, żeby rosły im pióra? Czy dzieci *uczą się* chodzić i mówić? Ta linia (jeśli w ogóle takowa istnieje) jest istotna dla naszych

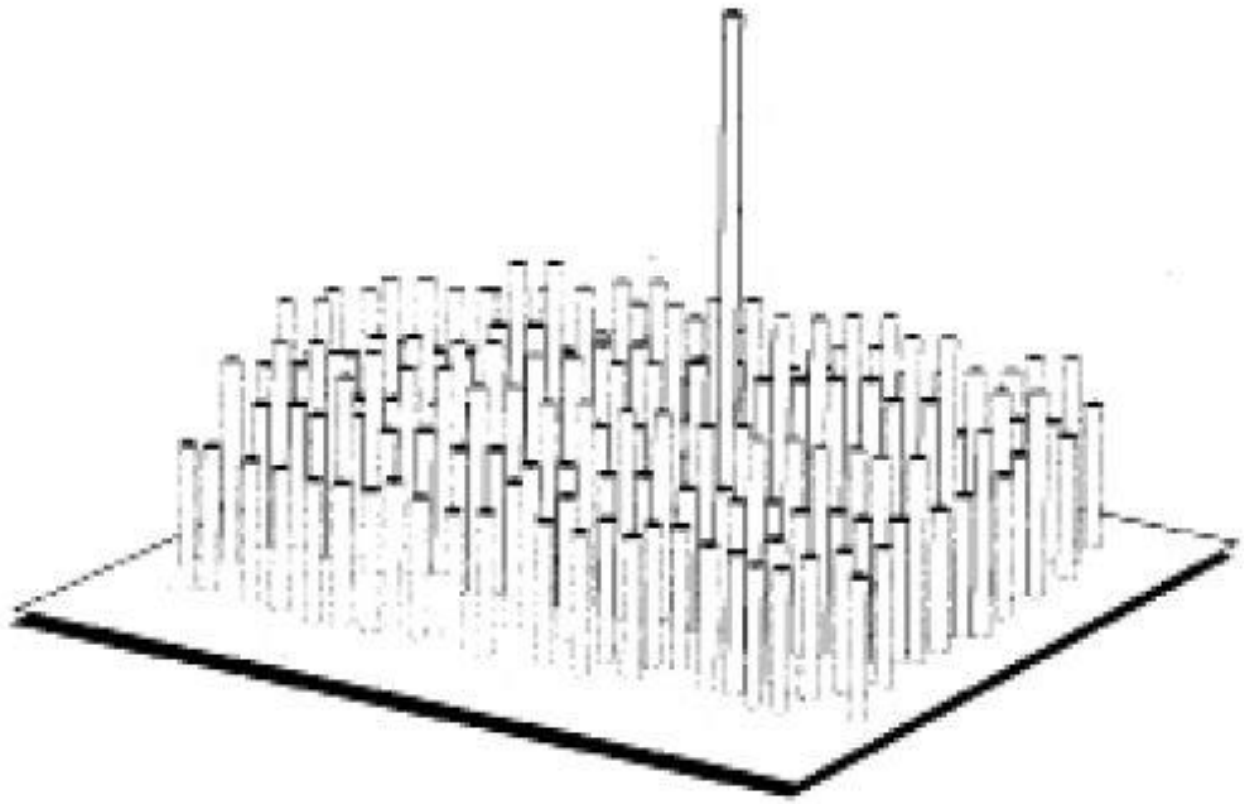
celów. Nazwijmy więc każdy taki proces, znajdujący się gdziekolwiek na spektrum od uczenia się ogniskowania oczu po uczenie się mechaniki kwantowej, *dopracowywaniem konstrukcji po urodzeniu*. Gdy się rodzisz, nadal pozostaje pewien obszar zmian, który w końcu zostaje ustalony tak czy inaczej jako względnie stały element konstrukcji do końca twojego życia (jeśli nauczysz się jeździć na rowerze albo mówić po rosyjsku, zwykle tego nie zapominasz).

Jak możliwy jest proces dopracowywania konstrukcji po urodzeniu? Tylko w jeden (nienadprzyrodzony) sposób: na drodze procesu bardzo podobnego do tego, który ustala prenatalną konstrukcję, albo, innymi słowy, podobnego do ewolucji przez dobór naturalny zachodzącej w jednym osobniku (w ramach fenotypu). Coś, co już znalazło się w jednostce na drodze zwykłego doboru naturalnego, musi odgrywać rolę mechanicznego selektora, a coś innego musi odgrywać rolę licznych przedmiotów doboru. Zaproponowano wiele różnych teorii działania tego procesu, jednak wszystkie – z wyjątkiem tych, które są po prostu wariackie lub zbyt tajemnicze – mają tę strukturę, a różnią się jedynie szczegółami proponowanych mechanizmów. W XX wieku najbardziej wpływową teorią długo był behawioryzm Barrhusa Frederica Skinnera, w którym pary bodziec-reakcja podlegały doborowi, a „wzmacniające" bodźce były mechanizmami doboru. Rola bodźców nagradzających i karzących – marchewki i kija – w kształtowaniu zachowania jest niezaprzeczalna, lecz behawiorystyczny mechanizm „warunkowania instrumentalnego" został powszechnie uznany za zbyt prosty, aby wyjaśniał złożoność dopracowywania konstrukcji po urodzeniu w gatunkach tak złożonych jak ludzie (i prawdopodobnie również u gołębi, ale to już inna historia). Obecnie nacisk kładzie się na różne teorie wyjaśniające proces ewolucji w mózgu (Dennett 1974). Różne wersje tego pomysłu są obecne od dekad, jednak dziś, gdy istnieje możliwość przetestowania rywalizujących ze sobą modeli w ogromnych symulacjach komputerowych, pojawiają się spory, od których mądrze jest trzymać się z daleka[56].

Wystarczy nam rzec, że w ten czy inny sposób plastyczny mózg jest w stanie adaptacyjnie się przeorganizować w odpowiedzi na określone nowości napotkane w środowisku, a proces, w wyniku którego tak się dzieje, jest niemal z pewnością mechaniczny i wyraźnie analogiczny do selekcji naturalnej. To pierwszy nowy ośrodek ewolucji: dopracowywanie konstrukcji po urodzeniu w indywidualnych mózgach. Doborowi podlegają różne struktury mózgowe regulujące czy wpływające na zachowanie, a dobór zachodzi na drodze takiego czy innego mechanicznego procesu odsiewania, który sam w sobie jest zainstalowany genetycznie w układzie nerwowym.

Niesamowite jest to, że ta zdolność, sama w sobie będąca wytworem ewolucji genetycznej przez dobór naturalny, nie tylko daje przewagę mającym ją organizmom nad ich zaprogramowanymi kuzynami, którzy nie mogą się przeprojektować, ale też odbija się na procesie ewolucji genetycznej i *przyspiesza go*. Jest to zjawisko mające wiele nazw, jednak najbardziej znana to „efekt Baldwina" (Richards 1987; Schull 1990). Przebiega ono w następujący sposób.

Wyobraźmy sobie populację pewnego gatunku, w którym występują duże różnice w momencie narodzenia w zakresie organizacji mózgów. Załóżmy, że jedna z możliwych organizacji wyposaża swojego nosiciela w pewną dobrą sztuczkę – talent behawioralny, który go ochrania lub zdecydowanie zwiększa jego szanse. Możemy to przedstawić w ramach tak zwanego krajobrazu adaptacyjnego; wysokość reprezentuje dostosowanie (im wyższe, tym lepiej), a długość i szerokość reprezentują zmienne oznaczające organizację mózgu (w tym eksperymencie myślowym nie musimy ich precyzować).



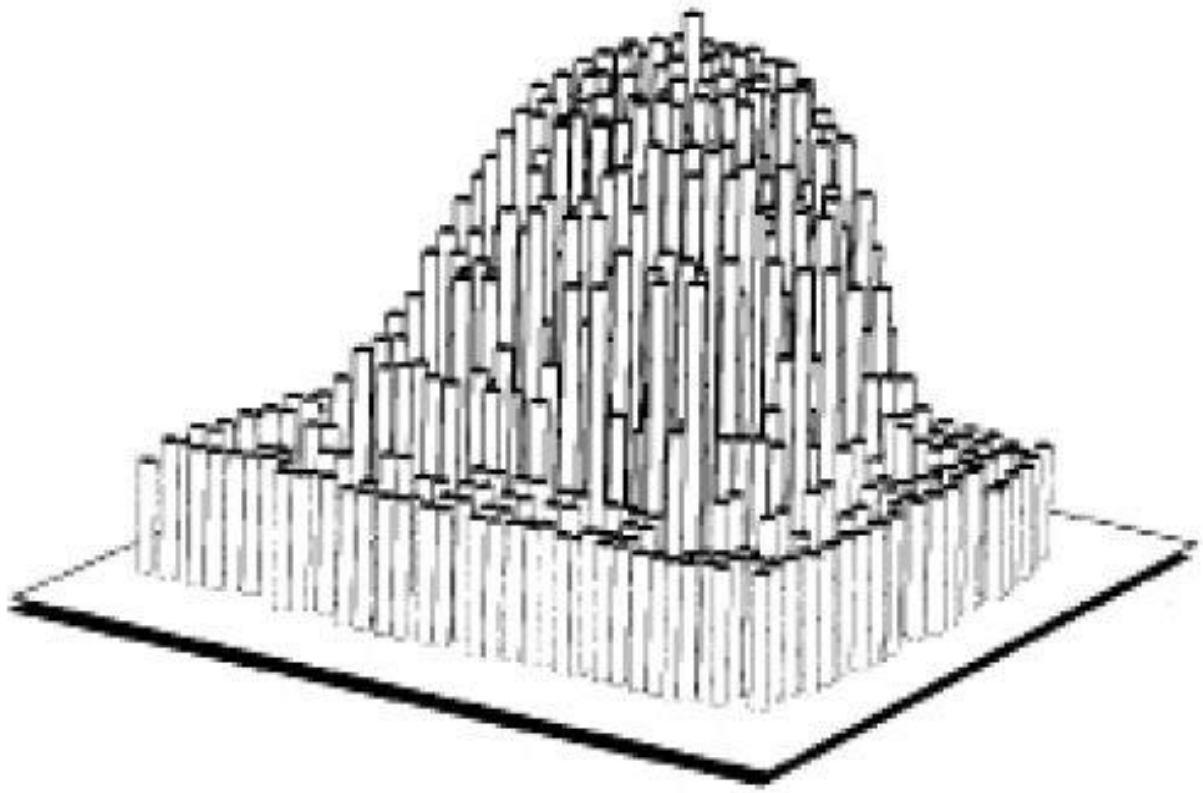
Ryc. 7.1

Jak pokazuje rycina, tylko jedna cecha jest preferowana; inne, bez względu na to, jak „blisko" preferowanej cechy się znajdują, mają takie samo dostosowanie (*fitness*). Taka igła w stogu siana może być praktycznie niewidoczna dla doboru naturalnego. Nawet jeśli kilku szczęśliwców ją ma, szanse, aby ich szczęście rozprzestrzeniło się w populacji przyszłych pokoleń, są zdecydowanie małe, *chyba że* wśród jednostek występuje plastyczność konstrukcji.

Załóżmy więc, że każda z tych jednostek ma inny genetyczny start, jednak w czasie swojego życia przemieszcza się w przestrzeni dostępnych im możliwości konstrukcyjnych dzięki swej plastyczności. Ze względu na pewne warunki panujące w środowisku wszystkie kierują się ku pożądanej cesze. Istnieje jedna dobra sztuczka, której można nauczyć się w ich środowisku, a wszystkie się jej uczą. Wyobraźmy sobie, że sztuczka jest tak dobra, że ci, którzy nigdy się jej nie uczą, mają ogromną wadę, i załóżmy, że przeżywają życie z konstrukcjami bardziej odsuniętymi w przestrzeni (i w związku z tym wymagają więcej przeprojektowania po urodzeniu) niż ci, którzy są bliscy dobrej sztuczki.

Pewna bajka (zaadaptowana z Hinton i Nowlan 1987) pomoże nam sobie to wyobrazić. Załóżmy, że istnieje dziesięć miejsc w mózgu każdego zwierzęcia, które mogą być zorganizowane na jeden z dwóch sposobów, A lub B. Załóżmy, że dobra sztuczka jest zaprojektowana jako AAABBBAAAA oraz że wszystkie inne organizacje są równie mało imponujące pod względem ich wpływu na zachowania. Skoro wszystkie te połączenia są plastyczne, każde zwierzę w czasie swojego życia może wypróbować którąkolwiek z 2^{10} różnych kombinacji cech A i B. Te zwierzęta, które rodzą się w stanie takim jak BAABBBAAAA, są oddalone o tylko jeden element od dobrej sztuczki (choć oczywiście mogą się od niej oddalić,

próbując najpierw szeregu innych elementów). Inne, które zaczynają od stanu BBBAAABBBB, muszą wykonać co najmniej 10 przeprojektowań (zakładając, że nie przeprojektują czegoś z powrotem, czyli błędnie), zanim odnajdą dobrą sztuczkę. Te zwierzęta, których mózgi zaczynają bliżej celu, będą miały wyższą przeżywalność od tych, które zaczynają daleko, nawet jeśli nie ma *innej* przewagi selekcyjnej, gdy jest się urodzonym ze strukturą znajdującą się „blisko" pożądanej kombinacji, nad strukturą „daleką" od kombinacji docelowej (jak wyraźnie pokazuje Ryc. 7.1). Zatem populacja w następnym pokoleniu będzie się składać głównie z jednostek bliższych celowi (które jednocześnie będą mogły łatwiej dojść do celu w czasie swojego życia), a procesy te mogą trwać do momentu, w którym cała populacje jest *genetycznie* nastawiona na dobrą sztuczkę. W ten sposób „odkryta" przez jednostki dobra sztuczka może być wówczas przekazywana dalej, względnie szybko, przyszłym pokoleniom.



Ryc. 7.2

Jeśli jednostki mają zmienną szansę dotarcia (a następnie „rozpoznania" oraz „przyczepienia się") do dobrej sztuczki w czasie swojego życia, niemalże niewidoczna igła w stogu siana na ryc. 7.1 stanie się szczytem o dość widocznym wybrzuszeniu, po którym może się wspinać dobór naturalny (Ryc. 7.2). Ten proces, nazywany „efektem Baldwina", może z początku wyglądać jak zdyskredytowany pomysł Lamarcka, mówiący o genetycznym przekazywaniu cech nabytych, ale nim nie jest. Nic, czego uczy się jednostka, nie jest przekazane jej potomstwu. Po prostu jednostki, które mają tyle szczęścia, żeby być bliżej dobrej sztuczki w przestrzeni konstrukcyjno-eksploracyjnej, będą zwykle miały więcej potomstwa, które również będzie bliżej dobrej sztuczki. Po upływie pokoleń rywalizacja stanie się bardziej zacięta: w końcu jedynymi osobnikami, które mogą ze sobą konkurować, są te, które urodziły się z dobrą sztuczką

(lub bardzo blisko niej). Jednak bez plastyczności efekt by nie istniał, gdyż „prawie robi wielką różnicę", *chyba że* próbujesz różnych układów, aż w końcu trafisz do celu.

Można powiedzieć, że dzięki efektowi Baldwina jednostki wstępnie testują skuteczność konkretnych konstrukcji przez fenotypową (indywidualną) eksplorację przestrzeni najbliższych możliwości. Jeśli w ten sposób zostaje odkryty szczególnie korzystny układ, owo odkrycie *stworzy* nową presję selekcyjną: organizmy bliższe w krajobrazie adaptacyjnym do tego odkrycia będą miały znaczną przewagę nad organizmami od niego oddalonymi. Oznacza to, że gatunki wyposażone w plastyczność będą miały *skłonność* do ewolucji szybszej (oraz „wnikliwszej") niż te gatunki, które się nią nie cechują. Zatem ewolucja w drugim ośrodku, plastyczności fenotypowej, może wzbogacić ewolucję w ośrodku pierwszym, czyli w różnorodności genetycznej. (Wkrótce przyjrzymy się efektowi równoważącemu, pojawiającemu się w wyniku interakcji z trzecim ośrodkiem).

4. Plastyczność w ludzkim mózgu: początki

Otóż tak samo rozum, korzystając z wrodzonych swoich sił, przysposabia sobie narzędzia umysłowe, dzięki którym zdobywa nowe siły do innych dzieł umysłowych, a z dzieł tych z kolei nowe ma narzędzia, czyli zdolność do dalszych badań, i tak coraz dalej się wspina, póki na szczycie mądrości nie stanie.

Baruch Spinoza, 1677/2009, s. 354

Układy nerwowe z całkowicie wrodzoną organizacją są lekkie, energetycznie wydajne oraz znakomite dla organizmów żyjących w środowiskach przewidywalnych i mających ograniczony budżet. Bardziej wymyślne mózgi dzięki swojej plastyczności są zdolne nie tylko do stereotypowego antycypowania, ale również do dostosowywania się do trendów. Nawet skromna ropucha ma w pewnym stopniu swobodę co do tego, jak zareaguje na nowość, powoli zmieniając wzorce zachowań, aby śledzić – po dłuższym upływie czasu – te zmiany w charakterystyce środowiska, które mają największy wpływ na jej dobrostan (Ewert 1987). W mózgu ropuchy konstrukcja zajmująca się światem ewoluuje wielokrotnie szybciej niż dobór naturalny – „pokolenia" trwają sekundy lub minuty, nie lata. Jednak żeby uzyskać sterowanie rzeczywiście dużej mocy, będziemy potrzebować urządzenia antycypującego, które dostosuje się bardzo w ciągu milisekund, a do tego będziemy potrzebować wirtuoza wytwarzania przyszłości, systemu umiejącego myśleć naprzód, unikającego rutyny w swoich działaniach, rozwiązującego problemy, zanim się one pojawią, i rozpoznającego zupełnie nowe zwiastuny dobra i zła. Pomimo naszej lekkomyślności my, istoty ludzkie, jesteśmy nieporównanie lepiej wyposażeni do tego zadania niż jakiekolwiek inne samosterujące gatunki, a dzieje się to za sprawą naszych ogromnych mózgów. Tylko jak?

Przyjrzyjmy się naszym postępom. Naszkicowaliśmy historię – pojedynczy wątek wielowymiarowej tkaniny historii ewolucyjnej – ewolucji mózgu naczelnych. Ta ewolucja opiera się na tysiącach lat historii wcześniejszych układów nerwowych, więc taki mózg składa się z mieszaniny wyspecjalizowanych obwodów o konstrukcji wykonującej określone zadania opłacalne dla przodków naczelnych: detektorów poruszających się przedmiotów połączonych z mechanizmami uchylania się, detektorów organizmów patrzących się na nas połączonych z urządzeniami odróżniającymi przyjaciela od wroga i od pożywienia, połączonych z odpowiednimi dalszymi mechanizmami. Możemy też dodać tak charakterystyczne dla naczelnych obwody, takie jak koordynacja ręka-oko, służąca do zbierania owoców leśnych czy nasion, czy też inne zaprojektowane do chwytania gałęzi bądź nawet radzenia sobie

z przedmiotami blisko twarzy lub pyska (Rizzolatti, Gentilucci i Matelli 1985). Dzięki ruchomym oczom oraz skłonności do eksploracji i uaktualniania informacji mózgi naczelnych regularnie odbierały morze informacji przez różne zmysły (to informacje wielomodalne, jak powiedziałby neuronaukowiec), a wówczas pojawił się nowy problem: problem kontroli na najwyższym poziomie.

Ten problem jest również szansą, bramą do nowego fragmentu przestrzeni konstrukcji. Zakładamy, że dotychczas układy nerwowe rozwiązywały problem „Co mam teraz zrobić?" przez względnie prosty akt zachowania równowagi między ściśle ograniczonym repertuarem czynności – jeśli nie słynnymi czterema F (walcz, uciekaj, zjedz lub kopuluj[57]), to ich skromnym rozwinięciem. Jednak wraz ze wzrostem funkcjonalnej plastyczności oraz wzrostem dostępności „scentralizowanych" informacji pochodzących od przeróżnych specjalistów problem tego, co zrobić teraz, zrodził metaproblem: o czym teraz *myśleć*. Można wyposażyć się w procedurę „wszystkie ręce na pokład", ale gdy już wszystkie ręce są na pokładzie, należy znaleźć jakiś sposób radzenia sobie z zalewem ochotników. Nie powinniśmy się spodziewać, że pod ręką już istnieje jakiś wygodny kapitan (co miałby robić dotąd?), więc konflikty pomiędzy ochotnikami muszą rozwiązać się bez żadnego wyższego kierownika. (Jak już widzieliśmy w przykładzie z układem immunologicznym, wspólnie podejmowane działania nie muszą zawsze zależeć od kontroli centralnego kierownictwa). Pionierski model tego rodzaju procesu w sztucznej inteligencji to architektura Pandemonium Olivera Selfridge'a (1959), w której wiele „demonów" równocześnie ubiega się o hegemonię, a ponieważ termin ukuty przez Selfridge'a na oznaczenie tego rodzaju architektury jest trafny, będę go używał w tej książce w sensie ogólnym na oznaczenie jego architektury oraz jej następców, bezpośrednich i pośrednich, takich jak „planowanie rywalizacyjne" (*contention scheduling*) (Norman i Shallice 1980; Shallice 1988) oraz sieci Zwycięzca Bierze Wszystko (ang. *Winner Take All*) Ballarda i Feldmana (1982) oraz ich następców.

Planowanie rywalizacyjne w stylu Pandemonium, kierowane dosyć bezpośrednio przez aktualne cechy środowiska, nadal daje układ nerwowy o ograniczonej możliwości uprzedzania przyszłości. Tak jak Odmar Neumann postawił hipotezę, że reakcje orientacyjne, początkowo wywoływane przez nowości w środowisku, zaczęły być następnie inicjowane endogenicznie (od wewnątrz), tak i my możemy postawić hipotezę, iż istniała presja, aby rozwinąć bardziej wewnętrzny sposób rozwiązania metaproblemu, o czym należy myśleć, presja, aby utworzyć coś wewnątrz – coś mające więcej możliwości organizacyjnych wyimaginowanego kapitana.

Pomyślmy, jak w tym momencie wyglądało zachowanie hipotetycznego przodka naczelnych z zewnątrz (odsuwamy na później wszelkie pytania dotyczące tego, jak to jest być takim naczelnym): zwierzę zdolne do nauczenia się nowych sztuczek oraz niemal zawsze czujne i wrażliwe na nowości, jednak z „małym zakresem uwagi" i tendencją do koncentrowania jej na rozpraszających elementach środowiska. Każdy długotrwały projekt przekraczałby możliwości tego zwierzęcia, przynajmniej nowy projekt. (Powinniśmy zostawić miejsce na stereotypowe długotrwałe procedury zainstalowane genetycznie, takie jak budowa gniazd u ptaków, budowa tam u bobrów czy łapanie pożywienia przez ptaki i wiewiórki).

Chcemy sobie teraz wyobrazić budowanie bardziej ludzkiego umysłu na podłożu takiego układu nerwowego, z czymś na kształt „strumienia świadomości" zapewniającego utrzymanie wyrafinowanego toku myślenia, od którego najwyraźniej zależy ludzka cywilizacja. Szympansy to nasi najbliżsi krewni – genetycznie bliższe są nam niż gorylom czy orangutanom – i obecnie uważa się, że mieliśmy wspólnego przodka z szympansami około sześciu milionów lat temu. Od tego czasu nasze mózgi dramatycznie się rozeszły, jednak głównie pod względem wielkości, a nie struktury. Szympansy mają mózgi z grubsza takiej samej wielkości jak nasz wspólny

przodek (i należy – choć to trudne – pamiętać o tym, że szympansy same wyewoluowały od naszego wspólnego przodka w innym kierunku), lecz mózgi naszych przodków, hominidów, powiększyły się czterokrotnie. Ten wzrost objętościowy nie nastąpił szybko; przez kilka milionów lat po rozdzieleniu z protoszympansami nasi przodkowie, hominidzi, radzili sobie jakoś z mózgami wielkości mózgów małp człekokształtnych, mimo że stali się istotami dwunożnymi przynajmniej trzy i pół miliona lat temu. Gdy zaczęły się epoki lodowcowe, jakieś dwa i pół miliona lat temu, rozpoczęła się wielka encefalizacja, która zakończyła się około 150 000 lat temu – *zanim* powstał język, gotowanie czy rolnictwo. Dlaczego mózgi naszych przodków urosły tak bardzo w tak krótkim czasie (w ewolucyjnej skali czasowej była to raczej eksplozja niż rozkwit), jest kwestią kontrowersyjną (książki Williama Calvina są źródłem pouczających informacji na ten temat). Bezsporna jest jednak natura wytworu tego procesu: mózg wczesnych *Hominum sapientium* (którzy żyli od około 150 000 lat temu do końca ostatniej epoki lodowcowej jedyne 10 000 lat temu) był bardzo złożony i cechował się ogromną plastycznością; był niemal nieodróżnialny co do wielkości i kształtu od naszego mózgu. Ważna jest rzecz następująca: ten spektakularny wzrost wielkości mózgu hominidów właściwie zakończył się, *zanim* rozwinął się język, więc nie może być odpowiedzią na trudności umysłowe, które powstały wraz z językiem. Noam Chomsky i wielu innych stawia hipotezę, że istnieją wewnętrzne specjalizacje w zakresie mowy, a hipoteza ta zaczyna obecnie być szczegółowo potwierdzana przez neuroanatomię. Specjalizacje te są jednak *bardzo* niedawnymi, pospiesznymi dodatkami, wykorzystującymi z pewnością wcześniejsze obwody służące do analizy sekwencji (Calvin 1989a), a przyspieszonymi przez efekt Baldwina. Co więcej, najbardziej niezwykły rozwój ludzkich możliwości *umysłowych* (których wytworami są rozwój gotowania, rolnictwa, sztuki i jednym słowem cywilizacji) nastąpił jeszcze później, od czasu końca ostatniej epoki lodowcowej, w ciągu 10 000 lat, czyli mrugnięcia, które z perspektywy ewolucyjnej jest niemalże natychmiastowe, gdyż trendy w ewolucji rozwijają się w ciągu milionów lat. Nasze mózgi w momencie urodzenia są wyposażone w niewiele umiejętności, jeśli w ogóle jakieś, których nie miały mózgi naszych przodków 10 000 lat temu. A zatem potężny rozwój u *Hominum sapientium* w ciągu ostatnich 10 000 lat musiał być prawie całkowicie spowodowany radykalnie nowym wykorzystaniem plastyczności ich mózgu – poprzez stworzenie pewnego rodzaju *oprogramowania*, zwiększającego jego ukryte siły (Dennett 1986).

Krótko mówiąc, nasi przodkowie musieli się nauczyć pewnych dobrych sztuczek, które byli w stanie wykonać za pomocą swojego dostrajanego sprzętu. Te sztuczki nasz gatunek dopiero zaczął zapisywać, poprzez efekt Baldwina, w genomie. Jak zresztą zobaczymy, istnieją powody, aby wierzyć, że mimo początkowej presji selekcyjnej, nakazującej stopniowe wrodzone instalowanie tych dobrych sztuczek, sztuczki te tak bardzo zmieniły naturę środowiska dla naszego gatunku, że nie istnieje już istotna presja selekcyjna nakazująca dalsze zmiany w konstrukcji mózgu. Jest prawdopodobne, że niemal cała selekcyjna presja wywierana na konstrukcję ludzkiego układu nerwowego została odparta przez skutki uboczne tych nowych możliwości konstrukcyjnych wykorzystanych przez naszych przodków.

Dotychczas starałem się unikać określania prostszych układów nerwowych jako *reprezentujących* cokolwiek w świecie. Różne omawiane przez nas konstrukcje, zarówno te plastyczne, jak i te wprowadzające wrodzone instalacje, można postrzegać jako wrażliwe na informacje o różnych cechach środowiska organizmu lub reagujące na nie, lub też zaprojektowane do korzystania z nich. Stąd w tym minimalnym sensie mogłyby być one nazwane reprezentacjami, jednak teraz powinniśmy się zatrzymać i pomyśleć, jakie cechy tak skomplikowanych konstrukcji powinny uzasadniać tezę, że są one systemami reprezentacji.

Pewna zmienność w mózgu jest po prostu potrzebna jako nośnik krótkotrwałych wzorców

aktywności mózgu, które w jakiś sposób *rejestrują* czy przynajmniej *śledzą* cechy w sposób istotny zmienne w środowisku. Coś w mózgu musi się zmienić, aby możliwe było śledzenie toru lotu ptaka, spadku temperatury powietrza czy jednego z wewnętrznych stanów samego organizmu – spadku poziomu cukru we krwi, wzrostu poziomu dwutlenku węgla w płucach. Oprócz tego – i jest to punkt podparcia, który daje dźwignię prawdziwej reprezentacji – te krótkotrwałe wewnętrzne wzorce mogą w końcu być w stanie kontynuować „śledzenie" (w szerokim sensie) cech, do których się odnoszą, gdy są czasowo odcięte od związków przyczynowych ze swoimi desygnatami. „Zebra, która zauważyła lwa, nie zapomina, gdzie się on znajduje, gdy na chwilę przestaje go obserwować. Lew nie zapomina, gdzie jest zebra" (Margolis 1987, s. 1987). Porównajmy to do prostszego zjawiska – słonecznik podąża za ruchem słońca po niebie, dostosowując swój kąt niczym ruchomy panel słoneczny, aby zmaksymalizować ilość światła słonecznego, które na niego pada. Jeśli słońce zostaje chwilowo zakryte, słonecznik nie może podążać za jego ruchem; mechanizm czuły na ruch słońca nie reprezentuje ruchu słońca w szerokim rozumieniu. Zaczątki prawdziwej reprezentacji można znaleźć u wielu niższych zwierząt (i nie powinniśmy wykluczać *a priori* możliwości prawdziwej reprezentacji u roślin), jednak u ludzi możliwość reprezentacji wzrosła w szalonym tempie.

Wśród rzeczy, które dorosły mózg może reprezentować, są nie tylko:

(1) położenie ciała i jego kończyn;

(2) plama czerwonego światła;

(3) stopień głodu;

(4) stopień pragnienia;

(5) zapach dobrego, starego, czerwonego burgunda;

(6) zapach dobrego, starego, czerwonego burgunda jako zapachu Chambertin, rocznik 1971;

(7) Paryż;

(8) Atlantyda;

(9) pierwiastek kwadratowy z największej liczby pierwszej mniejszej niż 20;

(10) koncepcja składanego korkociągu powlekanego niklem oraz wyciągacza zszywek.

Wydaje się niemal pewne, że mózg żadnego innego zwierzęcia nie potrafi reprezentować punktów 6–10 oraz, że wymagany jest istotny proces dostosowywania się mózgu niemowlaka, zanim którekolwiek z nich będzie mogło być w ogóle zarejestrowane lub reprezentowane. Natomiast pierwsze pięć punktów wskazuje to, co niemal każdy mózg potrafi reprezentować (w jakimś sensie) bez żadnego ćwiczenia.

Tak czy owak, sposób, w jaki mózg reprezentuje głód, musi różnić się fizycznie od tego, jak reprezentuje pragnienie – ponieważ musi sterować innym zachowaniem w zależności od tego, co jest reprezentowane. Z drugiej strony musi również istnieć różnica między reprezentowaniem przez konkretny dorosły mózg Paryża a reprezentowaniem przezeń Atlantydy, gdyż myślenie

o jednym nie jest myśleniem o drugim. Jak konkretny stan czy zdarzenie mózgowe może reprezentować jedną cechę świata, ale nie inną?[58] No i co sprawia, że pewien element mózgu reprezentuje to, co reprezentuje, *jak to się dzieje*, że reprezentuje to, co reprezentuje? Tutaj ponownie (obawiam się, że to zdanie stanie się nudne!) istnieje wiele możliwości otwartych przez procesy ewolucyjne: pewne elementy systemu reprezentacji mogą być – czy raczej muszą być (Dennett 1969) – wrodzone, a inne muszą być „wyuczone". Pewne kategorie, które są istotne dla życia (jak głód czy pragnienie), są niewątpliwie „dane" nam wraz z urodzeniem, inne zaś musimy rozwinąć sami[59].

Jak to robimy? Prawdopodobnie tworząc i wybierając wzorce czynności neuronalnych w korze mózgowej, ogromnej, pozwijanej masie, która sprawnie rozwinęła się w ludzkiej czaszce i obecnie całkowicie przykrywa znajdujący się niżej starszy mózg zwierzęcy. Zwykłe powiedzenie, że jest to ewolucyjny proces zachodzący przede wszystkim w korze, pozostawia zbyt wiele niewyjaśnionych tajemnic, a na tym poziomie złożoności i zaawansowania, nawet gdybyśmy wyjaśnili ten proces na poziomie synaps czy węzłów neuronów, bylibyśmy zadziwieni innymi aspektami tego, co musi się dziać. Jeśli w ogóle mamy to zrozumieć, najpierw musimy wspiąć się na bardziej ogólny i abstrakcyjny poziom. Gdy już pojmiemy ten proces w zarysie na wyższym poziomie, będziemy mogli pomyśleć o ponownym zejściu na bardziej mechaniczny poziom mózgu.

Plastyczność sprawia, że uczenie się jest możliwe, jednak jest jeszcze lepiej, gdy gdzieś w środowisku istnieje *coś do nauczenia*, co jest już wytworem wcześniejszego procesu konstruowania, aby każdy z nas nie musiał ponownie wynajdywać koła. Ewolucja kulturowa oraz przekazywanie jej wytworów to drugi nowy ośrodek ewolucji, a zależy on od fenotypowej plastyczności w taki sam sposób, jak fenotypowa plastyczność zależy od różnorodności genetycznej. Ludzie wykorzystują swoją plastyczność nie tylko po to, by się uczyć, lecz również po to, by uczyć się, jak się uczyć lepiej, a gdy już to osiągną, używają jej po to, by uczyć się lepiej, jak uczyć się lepiej, jak uczyć się lepiej i tak dalej. Wiemy też, jak udostępniać efekty uczenia nowicjuszom. W jakiś sposób *instalujemy* już wynaleziony i w dużej mierze „pozbawiony błędów" system nawyków w częściowo pozbawionym struktury mózgu.

5. Wynalezienie dobrych i złych nawyków autostymulacji

Jak mogę powiedzieć, co myślę,
zanim nie zobaczę, co mówię?
E.M. Forster, 1960

Mówimy nie tylko po ty, by powiedzieć innym, co myślimy, ale by powiedzieć sobie samym, co myślimy.
John Hughlings Jackson, 1915

Jak mogło dojść do takiego dzielenia się oprogramowaniem? Taka sobie bajeczka pokaże nam jedną z możliwości. Przenieśmy się do okresu w historii wczesnych *Hominum sapientium*, kiedy język – być może powinniśmy go nazwać „protojęzykiem" – dopiero zaczynał się rozwijać. Ci przodkowie byli dwunożnymi wszystkożercami, żyjącymi w małych grupkach spokrewnionych ze sobą osobników, i prawdopodobnie rozwinęli nawyki wokalizacyjne służące konkretnym celom, podobne jak u szympansów czy goryli, czy nawet u mniej z nami spokrewnionych gatunków jak koczkodany (Cheney i Seyfarth 1990). Możemy przypuszczać, że czynności komunikacyjne (czy też quasi-komunikacyjne) realizowane przez te wokalizacje nie

były czynnościami mowy w całym znaczeniu tego słowa (Bennett 1976), w którym intencja mówcy, jaką jest osiągnięcie pewnego efektu wśród publiczności, zależy od docenienia przez publiczność właśnie tej intencji[60]. Możemy jednak przypuszczać, że ci przodkowie, jak współczesne wokalizujące naczelne, potrafili rozróżnić różnych mówców i publiczności przy różnych okazjach, korzystając z informacji dotyczących tego, co obie strony uważały bądź chciały[61]. Na przykład hominid Alf nie zawracałby sobie głowy tym, by przekonać hominida Boba, że w jaskini nie ma jedzenia (burcząc „Niematujedzenia"), jeśli Alf wiedział, że Bob był już tego świadom. A jeśli Bob pomyślałby, że Alf chce go oszukać, byłby skłonny podejść do wokalizacji Alfa z ostrożnym sceptycyzmem[62].

Możemy spekulować, że czasem zdarzyło się, iż jeden z hominidów napotykał problemy i wtedy „prosił o pomoc", a szczególnie „prosił o informację". Czasem obecna publiczność odpowiadała poprzez „zakomunikowanie" czegoś, co miało odpowiedni wpływ na pytającego, pomagając mu spojrzeć na sprawę inaczej lub sprawiając, że „dostrzegł" rozwiązanie problemu. Aby ta praktyka zakorzeniła się w społeczności, pytający musieli być w stanie oddać przysługę i czasem wcielić się w rolę odpowiadającego. Musieli mieć behawiorystyczną możliwość bycia nakłonionym do przekazania „pomocnych" informacji, kiedy zostali o takowe „poproszeni" przez innych. Jeśli na przykład pewien hominid wiedział coś i został o to „spytany", mogło to mieć normalny efekt nakłaniający do „powiedzenia, co wie", jednak z pewnością mogły się od tego zdarzać wyjątki.

Ryc. 7.3

Innymi słowy, sugeruję, że w historii języka był moment, gdy wokalizacje służyły do uzyskiwania przydatnych informacji i dzielenia się nimi, ale nie możemy zakładać, że duch współpracy i wzajemnej pomocy miał znaczenie dla przetrwania lub że byłby systemem stabilnym, gdyby się pojawił. (Zob. np. Dawkins 1982/2003, s. 80; Sperber i Wilson 1986/2011). Zamiast tego musimy założyć, że koszty i zyski z uczestnictwa w takiej praktyce były w pewnym sensie „widoczne" dla tych istot i wiele z nich stwierdziło, że zyski *dla nich samych* przeważały nad kosztami, a zatem nawyki komunikacyjne na dobre zadomowiły się w tej społeczności.

Aż pewnego pięknego dnia (tej racjonalnej rekonstrukcji) jeden z tych hominidów „przypadkowo" poprosił o pomoc, kiedy w zasięgu słuchu nie było żadnej pomocnej publiczności – poza nim samym! Gdy usłyszał swoją własną prośbę, owa stymulacja wywołała

wytworzenie właśnie tego rodzaju wypowiedzi pomagającej innym, którą spowodowałaby prośba od innego osobnika. Ku swojej uciesze stworzenie stwierdziło, że właśnie samo siebie sprowokowało do odpowiedzi na własne pytanie.

Ten celowo uproszczony eksperyment myślowy ma na celu uzasadnienie stwierdzenia, że praktyka zadawania pytań sobie samemu mogła zostać zapoczątkowana jako naturalny skutek uboczny zadawania pytań innym, a jej użyteczność byłaby podobna: byłoby to zachowanie rozpoznane jako poprawiające czyjeś szanse, gdyż wspomagałoby działanie lepszymi informacjami. Aby ta praktyka była użyteczna, wystarczy, by pierwotnie istniejące relacje dostępu *wewnątrz* mózgu jednostki musiały być niezupełnie optymalne. Inaczej mówiąc, załóżmy, że odpowiednia informacja do jakiegoś celu już *jest w mózgu*, jednak pozostaje w rękach nie tego specjalisty; podsystem w mózgu potrzebujący informacji nie może uzyskać jej bezpośrednio od specjalisty – po prostu dlatego, że ewolucja nie stworzyła takiego „okablowania". Ale nakłanianie specjalisty do „nadawania" informacji do środowiska, a następnie wykorzystywanie istniejącej pary uszu (oraz układu słuchowego), która je usłyszy, byłoby sposobem na utworzenie „wirtualnego okablowania" między odpowiednimi podsystemami[63].

Taki akt autostymulacji mógł wyznaczyć nowy szlak między wewnętrznymi elementami jednostki. Mówiąc z grubsza, wepchnięcie pewnej informacji do czyichś uszu i układu słuchowego może równie dobrze być stymulacją dla tych rodzajów połączeń, których jednostka potrzebuje, może uruchomić prawidłowy mechanizm kojarzenia, umieścić odpowiedni umysłowy kąsek na końcu języka. Można więc to powiedzieć, usłyszeć siebie mówiącego właśnie to i w ten sposób otrzymać informację, której się oczekiwało.

Gdy prostackie nawyki wokalnej autostymulacji zaczęły rozprzestrzeniać się w formie dobrej sztuczki w zachowaniu populacji hominidów, można by się spodziewać, że szybko zostały udoskonalone zarówno w wyuczonych nawykach behawioralnych w populacji, jak i, dzięki efektowi Baldwina, w predyspozycjach genetycznych oraz dalszych ulepszeniach wydajności i efektywności. Możemy w szczególności spekulować, iż dostrzeżonoby większe walory mówienia do siebie półgłosem, co z kolei prowadziłoby do całkowitego ucichnięcia. Ten cichy proces podtrzymywałby pętlę autostymulacyjną, ale pozbyłby się peryferyjnej wokalizacji i słuchania, gdyż nie odgrywałyby one już żadnej roli. Ta innowacja miałaby później kolejny zysk, oportunistycznie zaakceptowany, polegający na prywatności kognitywnej stymulacji. (W następnym rozdziale rozważymy, jak te skrócone drogi komunikacji mogą działać). Taka prywatność byłaby szczególnie przydatna, gdyby rozumiejący osobnicy tego samego gatunku byli w zasięgu słuchu. To prywatne mówienie do siebie być może nie byłoby najlepszym wyobrażalnym ulepszeniem istniejącej funkcjonalnej architektury mózgu, jednak byłoby pod ręką, łatwe do odkrycia i z nawiązką by wystarczało. Byłoby powolne i żmudne w porównaniu z szybkimi, nieświadomymi procesami poznawczymi, na których się opierało, ponieważ musiałoby korzystać z długich szlaków układu nerwowego „zaprojektowanego do innych celów" – zwłaszcza do rozumienia słyszanej mowy. Byłoby tak samo linearne (ograniczone do jednego tematu naraz), jak społeczna komunikacja, z której wyewoluowało. I byłoby zależne, przynajmniej na początku, od kategorii informacyjnych ucieleśnionych w czynnościach, które wykorzystywało. (Jeśli istniało tylko pięćdziesiąt rzeczy, które jeden hominid mógł „powiedzieć" drugiemu, istniało tylko pięćdziesiąt rzeczy, które hominid mógł powiedzieć sobie).

Mówienie na głos jest jedną z możliwości. Rysowanie sobie obrazków jest innym aktem automanipulacji, który łatwo docenić. Załóżmy, że któregoś dnia jeden z tych hominidów bezmyślnie narysował dwie równoległe linie na dnie swojej jaskini, a gdy spojrzał na to, co zrobił, te dwie linie wizualnie przypomniały mu równoległe brzegi rzeki, którą musiał

przekroczyć jeszcze tego dnia, co z kolei przypomniało mu o tym, by zabrać swoją lianę, aby dostać się na drugą stronę. Możemy przypuszczać, że gdyby nie narysował „obrazka", poszedłby nad rzekę, a *następnie*, szybko się rozejrzawszy, zdałby sobie sprawę z tego, iż potrzebuje swojej liny, i musiałby po nią wrócić. Może to być zauważalna oszczędność czasu i energii, która mogłaby wytworzyć kolejny nawyk i w końcu udoskonalić się jako *prywatne* rysowanie diagramów „oczami umysłu".

Ludzki talent do wynajdywania nowych ścieżek wewnętrznej komunikacji czasem wyraźnie ujawnia się w przypadku uszkodzeń mózgu. Ludzie niezwykle często pokonują urazy mózgu, a *nigdy* nie jest to kwestia „uzdrowienia" czy naprawienia uszkodzonych obwodów. Ludzie raczej odkrywają nowe sposoby wykonywania starych czynności, a aktywna ekstrapolacja odgrywa dużą rolę w rehabilitacji. Szczególnie sugestywna anegdota pochodzi z badań nad pacjentami z rozszczepieniem mózgu (Gazzaniga 1978). Lewa i prawa półkula są zwykle połączone szerokim mostem włókien, zwanych „ciałem modzelowatym". Gdy zostanie ono chirurgicznie przecięte (przy leczeniu dotkliwej epilepsji), dwie półkule tracą swoje najważniejsze bezpośrednie „kable" łączące i praktycznie nie ma między nimi komunikacji. Gdy takiego pacjenta poprosimy o zidentyfikowanie przedmiotu – na przykład ołówka – przez włożenie ręki do torby i dotknięcie go, sukces zależy od tego, która ręka go dotknie. Większość okablowania w mózgu jest zorganizowana *kontralateralnie*, co oznacza, że lewa półkula dostaje informacje z prawej części ciała, którą również steruje i *vice versa*. Lewa półkula zwykle kontroluje język, dlatego gdy pacjent włoży do torby prawą rękę, może od razu powiedzieć, co w niej jest, lecz jeśli włoży do torby lewą rękę, tylko prawa półkula dostanie informację o tym, że przedmiotem jest ołówek, ale nie ma możliwości, aby to powiedzieć na głos. Jednak wydaje się, że czasami prawa półkula wpada na sprytny fortel: znajdując ostrą końcówkę ołówka i wbijając ją sobie w dłoń, pacjent sprawia, że ostry sygnał *bólu* zostaje wysłany do lewego ramienia, a niektóre włókna bólu działają *ipsilateralnie*. Lewa półkula, kontrolująca język, dostaje podpowiedź: jest to coś wystarczająco ostrego, aby sprawić ból. „Jest ostre – może to długopis? Ołówek?" Prawa półkula, słysząc tę wokalizację, może pomóc i dać podpowiedzi – zmarszczeniem brwi na dźwięk „długopis", uśmiechem na dźwięk „ołówek" – tak, że dzięki sprawnej grze w „dwadzieścia pytań" lewa półkula zostaje nakierowana na poprawną odpowiedź. Jest sporo anegdot o tak pomysłowych prowizorkach wynalezionych na poczekaniu przez pacjentów z rozszczepieniem mózgu, lecz powinniśmy podchodzić do nich ostrożnie. Być może są tym, czym się wydają: przypadkami pokazującymi zręczność, z jaką mózg może odkrywać i stosować strategie autostymulacyjne, aby ulepszyć wewnętrzną komunikację w przypadku nieobecności „pożądanego" okablowania. Mogą jednak również być nieświadomie upiększonymi fantazjami badaczy oczekujących takich świadectw. Na tym polega problem z anegdotami.

Moglibyśmy dalej się bawić, wymyślając kolejne wiarygodne scenariusze „wynalezienia" użytecznych trybów autostymulacji, jednak niosłoby to pewne ryzyko: można wtedy zapomnieć, że nie wszystkie tego rodzaju wynalazki muszą być użyteczne, aby przetrwać. Gdy ogólny nawyk autostymulacji eksploracyjnej został w jakiś sposób czy sposoby zaszczepiony, mogło się namnożyć mnóstwo niefunkcjonalnych (ale też niezbyt dysfunkcyjnych) wariacji. W końcu istnieje wiele rodzajów autostymulacji i automanipulacji, które przypuszczalnie w żaden istotny sposób nie wpływają na poznanie czy regulację zachowania, ale które z typowych darwinowskich powodów nie zanikły, a nawet mogą coraz bardziej utrwalać się (kulturalnie i genetycznie) w pewnych populacjach. Przypuszczalne przykłady to malowanie się na niebiesko, bicie się gałązkami brzozy, wycinanie wzorów na skórze, głodzenie się, powtarzanie sobie w kółko „magicznej" formuły czy wpatrywanie się w swój pępek. Jeśli te praktyki są nawykami wartymi wszczepienia, ich zalety jako przedmiotów zwiększających przystosowanie biologiczne są

przynajmniej niewystarczająco „oczywiste”, aby zostały wyniesione do rangi predyspozycji genetycznych, jednak może są zbyt świeże jak na tego rodzaju wynalazki.

Wiele różnych typów autostymulacji, poprawiających organizację poznawczą, jest teraz prawdopodobnie częściowo wrodzonych, a częściowo wyuczonych i indywidualnych. Tak jak można zobaczyć, że gładzenie siebie w pewien sposób może prowadzić do pewnych pożądanych skutków ubocznych, które są tylko częściowo i pośrednio kontrolowane – i można poświęcić wiele czasu i pomysłowości na rozwijanie i poznawanie technik produkowania tych skutków ubocznych – tak również można na wpół świadomie poznawać techniki autostymulacji poznawczej, rozwijając swój własny styl, mający określone mocne i słabe strony. Niektórzy są w tym lepsi od innych, a jeszcze inni nigdy nie nauczą się tych sztuczek, ale istnieje wiele sposobów dzielenia się i nauczania. *Przekaz* kulturowy, dopuszczając niemal każdego do dobrej sztuczki, może spłaszczyć szczyt wzgórza dostosowania (Ryc. 7.2), tworząc mały pagórek lub blat stołu, który zmniejsza presję selekcyjną przenoszenia sztuczki do genomu. Jeśli niemal wszyscy dają sobie radę na tyle, aby przetrwać w cywilizowanym świecie, presja selekcyjna przenoszenia dobrych sztuczek do genomu zostaje wygaszona, a przynajmniej umniejszona.

6. Trzeci proces ewolucyjny: memy i ewolucja kulturowa[64]

Tak jak nauczyliśmy się doić krowy, a potem udomowiliśmy je dla własnych korzyści, tak też nauczyliśmy się różnorodnie wykorzystywać umysły nasze i innych, a techniki stymulacji wzajemnej i autostymulacji głęboko zakorzeniły się w naszej kulturze i uczeniu. To, jak kultura stała się przechowalnią i ośrodkiem przenoszącym innowacje (nie tylko innowacje świadomości), jest ważne, aby zrozumieć źródła konstrukcji ludzkiej świadomości, gdyż jest ona kolejnym ośrodkiem ewolucji.

Jednym z pierwszych istotnych etapów w procesie samodzielnego i gruntownego konstruowania się ludzkiego mózgu po urodzeniu jest dostosowanie się do najistotniejszych warunków lokalnych: sprawnie (w ciągu dwóch lub trzech lat) staje się on mózgiem suahili, japońskim lub angielskim. Cóż to za przełom – niczym wejście w naciągniętą procę!

Jest dla nas bez znaczenia, czy proces ten nazywa się „uczeniem”, czy „rozwojem różnicowym”; odbywa się tak sprawnie i bez wysiłku, że jest niemal niewątpliwe, iż ludzki genom zawiera wiele adaptacji dostosowanych szczególnie do sprawnego nabywania języka. Wszystko to następuje jak na ewolucję bardzo szybko, jest to jednak dokładnie to, czego powinniśmy się spodziewać, znając efekt Baldwina. Umiejętność mówienia jest tak dobrą sztuczką, że każdy, kto jej nie opanował, był na przegranej pozycji. Nasz pierwszy mówiący przodek z pewnością musiał wykonać sporo pracy, aby opanować język, ale my jesteśmy potomkami ówczesnych wirtuozów[65].

Gdy nasze mózgi zbudowały już wejściowe i wyjściowe ścieżki dla nośników języka, szybko okazały się one *pełne pasożytów* (dosłownie, jak wkrótce zobaczymy), bytów, które wyewoluowały, aby prosperować we właśnie takiej niszy: *memów*. Schemat teorii ewolucji przez dobór naturalny jest jasny – ewolucja zachodzi wtedy, gdy zaistnieją następujące warunki:

(1) zmienność: nieprzerwana obfitość różnych elementów;

(2) dziedziczenie lub replikacja: elementy mogą tworzyć kopie lub się replikować;

(3) zróżnicowane „dostosowanie”: liczba kopii elementu tworzonych w danym przedziale czasowym zależy od interakcji między cechami tego elementu (cokolwiek, co sprawia, że są różne od innych elementów) a cechami środowiska, w którym trwa.

Zauważmy, że ta definicja, choć wzięta z biologii, nie mówi nic szczególnego o cząstkach organicznych, żywieniu czy nawet życiu. Jest to ogólna i abstrakcyjna charakterystyka ewolucji przez dobór naturalny. Jak stwierdził zoolog Richard Dawkins, najważniejsza zasada to:

[...] że wszelkie życie ewoluuje na drodze zróżnicowanej przeżywalności replikujących się bytów. [...] Przypadkiem, dominującą na naszej planecie replikującą się jednostką został gen, cząsteczka DNA. Ale przecież mogą być i inne. Jeśli tak, to o ile spełnione będą pewne warunki, jest niemal pewne, że staną się one podstawą procesów ewolucyjnych.

Czy jednak dla znalezienia innych rodzajów replikatorów, a tym samym innych rodzajów ewolucji musimy udawać się do odległych światów? Sądzę, że nowy rodzaj replikatora pojawił się właśnie na tej planecie. Co więcej, spotykamy się z nim twarzą w twarz. Wciąż jeszcze jest w powijakach, niezdarnie unosząc się w swoim bulionie pierwotnym, ale pod względem osiąganego tempa przemian ewolucyjnych zostawia stare, zdyszane geny daleko w tyle. [Dawkins 1976/1996, s. 266]

Te nowe replikatory to, z grubsza, idee. Nie „proste idee" Locke'a i Hume'a (idea koloru czerwonego czy idea okrągłości, ciepła lub zimna), ale pewnego rodzaju złożone idee, które stanowią osobne jednostki do zapamiętania – jak na przykład idea:

koła,
noszenia ubrań,
wendety,
trójkąta prostokątnego,
alfabetu,
kalendarza,
Odysei,
analizy matematycznej,
szachów,
rysowania w perspektywie,
ewolucji przez dobór naturalny,
impresjonizmu,
Greensleeves[66],
dekonstrukcjonizmu.

Intuicyjnie są to mniej lub bardziej rozpoznawalne jednostki kulturowe. Możemy jednak dokładniej powiedzieć, jak wyznaczamy granice – o tym, dlaczego *D-F#-A* nie jest jednostką, a motyw drugiej części VII symfonii Beethovena jest: jednostki są najmniejszymi elementami, które replikują się niezawodnie i w dużych ilościach. Dawkins tworzy pojęcie dla takich jednostek: *mem* –

jednostk[a] przekazu kulturowego, jednostk[a] *naśladownictwa*. Pasowałoby tu słowo „mimem", gdyż wywodzi się z odpowiedniego greckiego rdzenia. Mnie jednak potrzebne jest słowo monosylabowe, które choć trochę przypominałoby „gen" [...] słowo to można również uważać za spokrewnione z angielskim słowem *memory* lub francuskim słowem *même* (taki sam). [...]

Przykładami memów są melodie, idee, obiegowe zwroty, fasony ubrań, sposoby lepienia garnków czy budowania łuków. Tak jak geny rozprzestrzeniają się w puli genowej, przeskakując z ciała do ciała za pośrednictwem plemników lub jaj, tak memy propagują się w puli memów, przeskakując z jednego mózgu do drugiego w procesie szeroko rozumianego naśladownictwa. Jeśli naukowiec przeczyta czy usłyszy jakiś dobry pomysł, przekazuje go kolegom i studentom. Wspomina o nim w artykułach i na wykładach. O propagowaniu się nośnej idei można

powiedzieć wtedy, że przenosi się z mózgu do mózgu. [Dawkins 1976/1996, s. 266–267]

W *Samolubnym genie* Dawkins zachęca do dosłownego rozumienia idei ewolucji memów. Ewolucja memów nie jest tylko analogiczna do ewolucji biologicznej czy genetycznej, to nie tylko proces, który może zostać opisany metaforycznie w kategoriach ewolucyjnych, ale zjawisko podlegające właśnie prawom doboru naturalnego. Teoria ewolucji przez dobór naturalny jest neutralna, jeśli chodzi o różnice między genami i memami; są one po prostu różnymi rodzajami replikatorów ewoluujących w różnych ośrodkach i z różną prędkością. I tak jak geny zwierzęce nie mogły pojawić się na tej planecie, dopóki ewolucja roślin im tego nie umożliwiła (zapewniając atmosferę bogatą w tlen oraz gotowy zapas zamiennych substancji odżywczych), tak ewolucja memów nie mogła się rozpocząć, zanim umożliwiła im to ewolucja zwierząt, stwarzając *Homo sapiens* – z mózgami, które zapewniały schronienie, oraz z nawykami komunikacyjnymi zapewniającymi nośniki przekazu dla memów.

Jest to nowy sposób pojmowania idei. Mam nadzieję, że uda mi się również pokazać, iż jest to sposób dobry, jednak początkowo perspektywa, jaką proponuje, jest zdecydowanie niepokojąca, a nawet przerażająca. Możemy ją podsumować w sloganie: „Uczony to tylko metoda biblioteki służąca do tworzenia kolejnej biblioteki".

Nie wiem jak wy, ale ja na pierwszy rzut oka nie jestem zachwycony pomysłem, że mój mózg jest swego rodzaju kupą gnoju, w której odnawiają się larwy idei innych ludzi, zanim wyślą kopie siebie samych do informacyjnej diaspory. Wydaje się, że okrada to mój umysł ze znaczenia jako autora i jako krytyka. Zgodnie z tą wizją, kto rządzi – my czy nasze memy?

Nie ma oczywiście prostej odpowiedzi, a fakt ten leży u podłoża nieporozumień związanych z ideą *jaźni*. Ludzka świadomość jest w dużym stopniu wytworem nie tylko doboru naturalnego, ale również ewolucji kulturowej. Wkład memów w tworzenie naszych umysłów najłatwiej dostrzec, szczegółowo śledząc standardowe kroki myślenia ewolucyjnego.

Pierwszą zasadą dla memów, tak jak dla genów, jest to, że replikacja niekoniecznie istnieje ku pożytkowi czegoś; największe powodzenie mają replikatory, które są dobre w... replikowaniu! – jakikolwiek jest tego powód. Jak powiada Dawkins:

Mem, który powoduje, że noszące go ciała skaczą w przepaść, ma taką samą szansę przeżycia jak gen, który sprawia, że noszące go ciała skaczą w przepaść. Będzie eliminowany z puli memów [...]. Ale nie znaczy to przecież, że ostateczną miarą sukcesu w doborze memów jest przeżycie genów [...]. Oczywiście memy, który sprawiają, że osobniki go noszące zabijają się, działają niekorzystnie dla siebie, ale niekoniecznie śmiertelnie niekorzystnie [...]. Mem samobójstwa może się rozprzestrzeniać na przykład poprzez dramatyczne i dobrze nagłośnione męczeństwo, które zachęca innych do umierania „za sprawę", a ich śmierć z kolei inspiruje dalszych męczenników, i tak dalej. [Dawkins 1982/2003, s. 147–148]

Rzecz w tym, że nie ma *koniecznego* związku między mocą memu do powielania, jego dostosowaniem z *jego* punktu widzenia a jego wkładem do *naszego* dostosowania (bez względu na to, jak to oceniamy). Sytuacja ta nie jest zupełnie beznadziejna. Niektóre memy zdecydowanie manipulują nami, abyśmy brali udział w ich replikacji, *mimo* że uważamy je za niepotrzebne, ohydne czy nawet niebezpieczne dla naszego zdrowia i dobrobytu, to jednak wiele – większość, jeśli mamy szczęście – memów replikuje się nie tylko z naszym błogosławieństwem, ale *z powodu* naszego szacunku do nich. Myślę, że właściwie bezsporne jest to, iż niektóre memy są, koniec końców, dobre *z naszej perspektywy*, a nie tylko z ich perspektywy jako egoistycznych samoreplikatorów: takie ogólne memy jak współpraca, muzyka, pismo, edukacja, świadomość środowiska, ograniczenie użycia broni; oraz poszczególne memy jak: *Wesele Figara*, *Moby Dick*, butelki z kaucją, układ o ograniczeniu zbrojeń strategicznych. Inne memy są bardziej kontrowersyjne; wiemy, dlaczego się rozprzestrzeniają oraz dlaczego, koniec końców,

powinniśmy je tolerować pomimo problemów, które nam sprawiają: centra handlowe, fast foody, reklama w telewizji. Jeszcze inne są bezdyskusyjnie zgubne, ale niezwykle trudne do wykorzenienia: antysemityzm, porywacze samolotów, wirusy komputerowe, graffiti wykonywane sprayem.

Geny są niewidzialne; przenoszone są przez nośniki genów (organizmy), w których zwykle wywołują charakterystyczne skutki (efekty „fenotypowe"), wyznaczające ich los na dłuższą metę. Memy również są niewidzialne i przenoszone przez nośniki memów – obrazki, książki, powiedzenia (w poszczególnych językach, w formie ustnej i pisemnej, na papierze czy magnetycznie zakodowane itd.). Narzędzia i budynki oraz inne wynalazki to także nośniki memów. Wóz z kołami szprychowymi przewozi z miejsca na miejsce nie tylko towar; przewozi też świetny pomysł wozu z kołami szprychowymi z umysłu do umysłu. Istnienie memu zależy od fizycznego ucieleśnienia w jakimś ośrodku; jeśli wszystkie takie fizyczne ucieleśnienia zostają zniszczone, mem wyginie. Może oczywiście pojawić się niezależnie w późniejszym czasie – tak jak zasadniczo geny dinozaurów mogłyby ponownie powstać w jakiejś odległej przeszłości – jednak dinozaury, które by tak powstały i nosiły te geny, nie byłyby potomkami pierwotnych dinozaurów, a przynajmniej nie w sposób bardziej bezpośredni, niż jesteśmy nimi my. Los memów – to, czy liczne kopie ich kopii przetrwają i się rozmnożą – zależy od selektywnych sił oddziałujących bezpośrednio na fizyczne nośniki, które je ucieleśniają.

Nośniki memów zamieszkują nasz świat na równi z całą fauną i florą, małą i dużą. Ale w zasadzie są „widoczne" jedynie dla gatunku ludzkiego. Weźmy środowisko przeciętnego nowojorskiego gołębia, którego oczy i uszy codziennie atakuje mniej więcej tyle słów, obrazków i innych znaków czy symboli, przez ile atakowany jest ludzki mieszkaniec Nowego Jorku. Te fizyczne nośniki memów mogą istotnie rzutować na dobrostan gołębia, lecz nie przez memy, które ze sobą niosą – dla gołębia nie ma znaczenia, czy znaleziony przez niego okruszek leży pod „National Enquirer", czy pod „New York Timesem".

Dla ludzi jednak każdy nośnik memu jest potencjalnym przyjacielem lub wrogiem, niosącym ze sobą dar, który wzmocni nasze siły lub je osłabi niczym koń trojański, rozpraszający uwagę, obciążający pamięć, zakłócający osądy. Tych unoszących się w powietrzu najeźdźców można porównać do pasożytów, które dostają się do naszego ciała innymi drogami: istnieją korzystne pasożyty, takie jak bakterie w naszym układzie trawiennym, bez których nie moglibyśmy trawić jedzenia, znośne pasożyty, niewarte zachodu związanego z ich ewentualną eliminacją (na przykład niezliczona ich ilość na skórze i głowie), oraz zgubni najeźdźcy trudni do wyeliminowania (na przykład wirus HIV).

Jak na razie perspektywa memu może nadal wydawać się po prostu obrazowym ujęciem bardzo znanych obserwacji na temat wpływu elementów kultury na nas i na innych. Jednak Dawkins twierdzi, że w naszych wyjaśnieniach zwykle pomijamy zasadniczy fakt, iż „jakaś cecha kulturowa mogła powstać w takim, a nie innym kształcie po prostu dlatego, że jest to *korzystne dla niej samej*" (Dawkins 1976/1996, s. 276). To klucz do odpowiedzi na pytanie, czy powinniśmy wykorzystać i powielać mem memu. Zwykle uważa się, że następujące stwierdzenia są w zasadzie tautologiczne:

Ludzie wierzyli w ideę *X*, gdyż *X* było uważane za prawdziwe. Ludzie akceptowali *X*, gdyż uważali *X* za piękne.

Specjalnego wyjaśnienia wymagają zaś przypadki, w których piękna lub prawdziwa idea *nie jest* akceptowana albo *jest akceptowana*, mimo że jest ohydna i fałszywa. Natomiast z konkurencyjnego punktu widzenia, odwołującego się do pojęcia memu, odchylenia te można wyjaśnić w ogólny sposób – dla memu tautologią jest stwierdzenie, że:

Mem *X* rozprzestrzenił się wśród ludzi, gdyż *X* był dobrym replikatorem.

Istnieje nieprzypadkowa korelacja między tymi dwiema tautologiami; nie jest to zbieg okoliczności. Nie przetrwalibyśmy, gdybyśmy nie mieli nawyku, występującego częściej, niż wskazywałby tylko przypadek, wybierania memów, które nam pomagają. Nasze memowe układy immunologiczne nie są niezawodne, nie są też jednak beznadziejne. W praktyce możemy, z grubsza rzecz biorąc, liczyć na zgranie obu perspektyw: dobre memy to zwykle też te, które są również dobrymi replikatorami.

Teoria ta staje się interesująca tylko wtedy, gdy przyglądamy się wyjątkom, warunkom, w których obie perspektywy się rozchodzą; tylko jeśli teoria memów pozwoli nam lepiej zrozumieć te odchylenia od normalnego schematu, będzie miała jakąkolwiek szansę, by zostać przyjęta. (Zwróćmy uwagę, że zgodnie z koncepcją memu to, czy mem replikuje się skutecznie, jest ściśle niezależne od jego waloru epistemologicznego; może się on rozprzestrzenić pomimo swojej szkodliwości lub wyginąć pomimo swoich zalet).

Obecnie memy rozprzestrzeniają się po świecie z prędkością światła i replikują się w tempie, które sprawia, że nawet muszki owocówki i komórki drożdży wyglądają w porównaniu z nimi jak zastygłe w bezruchu. Bez zahamowań przenoszą się z jednego nośnika do drugiego, z jednego ośrodka do drugiego i pokazują, że są praktycznie nie do powstrzymania. Memy, jak geny, są *potencjalnie* nieśmiertelne, jednak, jak geny, zależą od istnienia ciągłego łańcucha nośników fizycznych, trwających mimo drugiej zasady termodynamiki. Książki są względnie trwałe, a inskrypcje na budynkach jeszcze trwalsze, lecz jeśli nie pozostają pod ochroną ludzkich konserwatorów, z czasem zanikają. Tak jak w przypadku genów, nieśmiertelność jest raczej kwestią replikacji niż trwałości pojedynczych nośników. Zachowanie memów platońskich przez serię kopii kopii jest szczególnie uderzającym przypadkiem. Choć niedawno odkryto niektóre fragmenty papirusu z tekstami Platona, powstałe z grubsza w czasach mu współczesnych, przetrwanie tych memów nie ma właściwie nic wspólnego z przetrwaniem na tak długą skalę. Dzisiejsze biblioteki zawierają tysiące, jeśli nie miliony fizycznych kopii (oraz tłumaczeń) *Menona* Platona, a kluczowi przodkowie transmisji tego tekstu zamienili się wieki temu w pył.

Sama fizyczna replikacja nośników nie wystarczy do zapewnienia memom długowieczności. Kilka tysięcy egzemplarzy nowej książki w twardej oprawie może zniknąć niemalże bez śladu w ciągu kilku lat, a kto wie, ile genialnych listów do redakcji, powielonych w setkach tysięcy kopii, znika na wysypiskach i w spalarniach śmieci każdego dnia? Może nastać dzień, kiedy to ewaluator memów niebędący człowiekiem będzie wybierać poszczególne memy mające przetrwać, jednak obecnie istnienie memów nadal zależy przynajmniej pośrednio od tego, czy co najmniej jeden ich nośnik przetrwał choćby krótki etap poczwarki w szczególnego rodzaju gnieździe memów: ludzkim umyśle.

Liczba umysłów jest ograniczona, a każdy z nich ma ograniczoną pojemność, istnieje więc znaczące współzawodnictwo między memami o to, aby dostać się do tylu umysłów, ile to możliwe. To współzawodnictwo jest główną siłą selekcyjną w memosferze, i tak jak w biosferze rozwiązania tego problemu okazały się wspaniale pomysłowe. Na przykład, jakiekolwiek zalety (z naszej perspektywy) mają poniższe memy, ich cechą wspólną jest to, iż mają fenotypowe przejawy, które na ogół sprawiają, że ich replikacja jest bardziej prawdopodobna, gdyż blokują lub uprzedzają siły środowiskowe mogące je wyeliminować: mem *wiary*, który hamuje zadanie swego rodzaju krytycznej oceny, mogącej zdecydować o tym, że idea wiary jest tak naprawdę pomysłem niebezpiecznym (Dawkins 1976/1996, s. 274); mem *tolerancji* i *wolności słowa*; mem włączania w list-łańcuszek ostrzeżenia o strasznych wypadkach losowych, które przydarzyły się osobom zrywającym łańcuszki w przeszłości; mem *teorii spiskowej*, który ma wbudowaną odpowiedź na zarzut, że nie ma silnego dowodu na spisek: „Oczywiście, że nie ma – właśnie tak potężny jest spisek!”. Niektóre memy są być może „dobre”, a inne „złe”; jednak wspólny jest dla

nich efekt fenotypowy, który systematycznie unieszkodliwia siły selekcyjne działające przeciwko nim. Memetyka populacyjna przewiduje, że w warunkach normalnych memy teorii spiskowej przetrwają niezależnie od swojej prawdy i że mem wiary jest w stanie zapewnić przetrwanie sobie oraz memom religijnym, które za sobą niesie, nawet w najbardziej racjonalistycznych środowiskach. Mem wiary rzeczywiście pokazuje *dostosowanie zależne od częstotliwości*: najlepiej rozkwita, gdy liczebnie przewyższają go memy racjonalistyczne; w środowisku z niewieloma sceptykami mem wiary zaczyna zanikać.

Inne pojęcia genetyki populacyjnej można przenieść równie łatwo: jest tu przypadek, który genetyk nazwałby „*sprzężone loci*": dwa memy, które są połączone ze sobą fizycznie akurat w taki sposób, że zwykle razem się replikują, co wpływa na ich szanse. Istnieje wspaniały marsz ceremonialny, znany wielu z nas, który byłby niezwykle często grany podczas ceremonii rozpoczęcia ślubów i przy innych świątecznych okazjach, skazując tym samym *Pomp and Circumstance* i *Marsz weselny* na wymarcie, gdyby nie fakt, że jego mem muzyczny jest zbyt ściśle związany z memem tytułowym, o którym wielu z nas ma skłonność myśleć, gdy tylko słyszymy muzykę: nienadające się do użytku dzieło sztuki sir Arthura Sullivana *Behold the Lord High Executioner!* (Oto starszy kat!)[67].

Ludzki umysł to przystań, do której przybyć pragną wszystkie memy, jednak jest on sam w sobie wytworem memów restrukturyzujących go, by stworzyć dla siebie lepsze siedlisko. Drogi wejścia i wyjścia są dopasowywane do warunków lokalnych i wzmacniane różnymi sztucznymi urządzeniami, które poprawiają wierność i długotrwałość replikacji: umysły rdzennych Chińczyków różnią się znacznie od umysłów rdzennych Francuzów, a umysły piśmienne dramatycznie różnią się od umysłów analfabetów. Memy w zamian za możliwość przebywania w ludzkich ciałach zapewniają mnóstwo korzyści – bez wątpienia wraz z jakimiś końmi trojańskimi dorzuconymi gratis. Zwykłe ludzkie mózgi nie są do siebie zupełnie podobne; znacząco różnią się rozmiarem, kształtem oraz ogromną liczbą szczegółów połączeń, od których zależy ich sprawność. Ale najbardziej zaskakujące różnice w zakresie ludzkiej sprawności zależą od mikrostrukturalnych różnic spowodowanych różnymi memami, które do nich weszły i tam zamieszkały. Memy wzmacniają okazje dla siebie nawzajem: na przykład mem edukacji to mem, który umacnia sam proces implantacji memów.

Jeśli jednak prawdą jest, że umysły ludzkie w ogromnym stopniu same są dziełem memów, nie możemy wówczas podtrzymać biegunowości spojrzenia, towarzyszącego wcześniej naszym rozważaniom; nie ma mowy o żadnym „memy przeciwko nam", ponieważ wcześniejsze infekcje memów zadecydowały o tym, *kim lub czym jesteśmy*. „Niezależny" umysł walczący o to, by chronić się przed obcymi i niebezpiecznymi memami, jest mitem; na niższym poziomie utrzymuje się napięcie między biologicznym imperatywem genów a imperatywem memów, lecz bylibyśmy niemądrzy, „stając po stronie" naszych genów – byłby to błąd socjobiologii w wersji pop. Na jakich zatem podwalinach możemy stanąć, gdy walczymy, aby twardo stąpać po ziemi w burzy memowej rozpętanej wokół nas? Jeśli moc replikacyjna nie decyduje o słuszności, jaki ma być wieczny ideał wyznaczający wartość memów dla „nas"? Powinniśmy wspomnieć, że memy koncepcji normatywnych – powinności, dobra, prawdy i piękna – są jednymi z najbardziej umocnionych mieszkańców naszych umysłów oraz że wśród memów, które nas konstytuują, odgrywają one główną rolę. Nasze istnienie nas jako nas, jako tego, czym jesteśmy jako osoby myślące – nie tego, czym jesteśmy jako organizmy – nie jest niezależne od tych memów.

Podsumowując: ewolucja memów ma potencjał, aby głęboko ulepszyć konstrukcję maszynerii leżącej u podstaw naszego mózgu – i to bardzo szybko w porównaniu z wolnym tempem genetycznych prac badawczo-rozwojowych. Zdyskredytowana idea Lamarcka dotycząca genetycznej transmisji cech nabytych indywidualnie była z początku atrakcyjna dla biologów

częściowo ze względu na to, że rzekomo ułatwiała wprowadzanie nowych wynalazków do genomu. (Bardzo dobra krytyka lamarkizmu znajduje się w książce Dawkinsa *Rozszerzony fenotyp*, 1982/2003). Coś takiego się nie dzieje ani nie mogłoby się wydarzyć. Efekt Baldwina przyspiesza ewolucję, sprzyjając przenoszeniu indywidualnie odkrytych dobrych sztuczek do genomu za pośrednictwem nowej presji selekcyjnej, czego rezultatem jest powszechne zaakceptowanie dobrych sztuczek przez jednostki. Jednak ewolucja kulturowa, która przebiega jeszcze szybciej, pozwala jednostkom nabyć poprzez transmisję kulturową dobre sztuczki udoskonalone przez poprzedników, którzy nie są nawet ich przodkami genetycznymi. Efekty takiego dzielenia się dobrymi konstrukcjami są tak silne, że ewolucja kulturowa prawdopodobnie zatarła niemal wszystkie delikatne presje efektu Baldwina. Poprawki w konstrukcji, jakie jednostka uzyskuje od swojej kultury – rzadko kiedy musimy „ponownie wynaleźć koło" – prawdopodobnie przeważają nad większością indywidualnych różnic genetycznych w konstrukcji mózgu, zabierając przewagę tym, którzy są odrobinę lepiej przystosowani zaraz po urodzeniu.

Wszystkie trzy ośrodki – ewolucja genetyczna, plastyczność fenotypowa i ewolucja memetyczna – przyczyniły się kolejno do konstrukcji ludzkiej świadomości, i to ze wzrastającą prędkością [ewolucji]. W porównaniu z plastycznością fenotypową, która istnieje od milionów lat, ewolucja memetyczna na większą skalę jest zjawiskiem nadzwyczaj świeżym, gdyż stała się istotną siłą dopiero w ciągu ostatnich setek tysięcy lat i przyczyniła się do burzliwego rozwoju cywilizacji mniej niż 10 000 lat temu. Ogranicza się do jednego gatunku, *Homo sapiens*, i możemy zauważyć, że to ona doprowadziła do pojawienia się kolejnego, czwartego ośrodka potencjalnych prac badawczo-rozwojowych dzięki memom nauki: jest nim bezpośrednie korygowanie indywidualnych układów nerwowych poprzez inżynierię neuronaukową oraz korygowanie genomu przez inżynierię genetyczną.

7. Memy świadomości: wirtualna maszyna do zainstalowania

Choć pewien organ mógł nie ukształtować się pierwotnie uformowany do jakiegoś konkretnego celu, to jeśli teraz mu służy, zasadne jest stwierdzenie, że powstał specjalnie w tym celu. Na tej samej zasadzie, gdyby ktoś miał skonstruować maszynę w jakimś konkretnym celu, ale używając starych kół, szprych i krążków, jedynie trochę zmienionych, o całej maszynie, ze wszystkimi częściami, można by powiedzieć, że została skonstruowana w tym celu. Przeto w całej przyrodzie niemal każda część każdego żyjącego stworzenia prawdopodobnie służyła, w stanie trochę zmienionym, innym celom oraz działała w żywej maszynerii wielu dawnych i przeróżnych form.

Charles Darwin, 1874

Duży mózg, podobnie jak duży rząd, może nie być w stanie robić prostych rzeczy w prosty sposób.

Donald Hebb, 1958

Najpotężniejszym czynnikiem procesu wznoszenia się człowieka jest radość, którą sprawiają mu własne umiejętności. Lubi robić to, co robi dobrze, a zrobiwszy coś dobrze, chciałby to wykonać jeszcze lepiej.

Jacob Bronowski, 1973/1988 (s. 115)

Elementem mojej spekulatywnej historii było to, że nasi przodkowie, tak jak my, czerpali przyjemność z różnego rodzaju względnie nieukierunkowanej samoeksploracji – ciągłego

stymulowania samych siebie, aby zobaczyć, co się stanie. Ze względu na plastyczność mózgu oraz wrodzony niepokój i ciekawość, dzięki którym badamy każdy zakątek naszego otoczenia (którego tak ważnym i wszechobecnym składnikiem są nasze własne ciała), z perspektywy czasu nie powinno nas dziwić, że odkryliśmy strategie autostymulacji czy automanipulacji, doprowadzające do wszczepienia nawyków i dyspozycji diametralnie zmieniających wewnętrzną strukturę komunikacyjną naszych mózgów, ani że te odkrycia stały się częścią kultury – memami – które następnie okazały się dostępne dla wszystkich.

Przekształcenie ludzkiego mózgu przez infekcje memów zmienia drastycznie rolę tego organu. Jak wspomnieliśmy, różnice między mózgami, z których jeden posługuje się językiem ojczystym Chińczyka, a drugi – angielskim, mogą wyjaśnić ogromne rozbieżności w kompetencji tych mózgów, natychmiast rozpoznawane w zachowaniu oraz istotne w wielu kontekstach eksperymentalnych. Przypomnijmy na przykład, jak ważne w eksperymentach z ludźmi dla badacza (heterofenomenologa) jest wiedzieć, czy badany zrozumiał instrukcje. Te różnice funkcjonalne, choć przypuszczalnie ucieleśnione we wzorcach mikroskopijnych zmian w mózgu, są dla neuronaukowców właściwie niewidoczne teraz, a zapewne już zawsze, więc jeśli mamy w ogóle zrozumieć funkcjonalną architekturę *wytworzoną* przez takie zakażenia memami, będziemy musieli znaleźć wyższy poziom, na którym to opiszemy. Na szczęście jeden jest dostępny, zapożyczony z informatyki. Potrzebny poziom opisu i wyjaśnienia jest *analogiczny do* (choć nie identyczny z nim) jednego z „poziomów oprogramowania" w opisie komputerów: to, co musimy zrozumieć, to jak ludzka świadomość może być realizowana przez działania *maszyny wirtualnej*, wytworzonej przez memy w mózgu.

Oto hipoteza, której będę bronił:

Ludzka świadomość jest *sama w sobie* ogromnym kompleksem memów (lub, ściślej rzecz biorąc, efektów memowych w mózgach), który może zostać najlepiej wyjaśniony jako działanie maszyny wirtualnej *w stylu von Neumanna zaimplementowanej* w *równoległej architekturze* mózgu, niezaprojektowanej do żadnej aktywności tego typu. Moce tej *maszyny wirtualnej* wzmacniają znacznie podstawowe siły organicznego *sprzętu*, na którym się ona opiera, ale jednocześnie wiele jej najbardziej osobliwych cech, a w szczególności jej ograniczenia, mogą być wyjaśnione jako skutki uboczne *prowizorek*, które umożliwiły owo dziwne, ale efektywne, ponowne wykorzystanie istniejącego organu do nowych celów.

Ta hipoteza wkrótce wyłoni się z gąszczu żargonu, w którym właśnie ją wyłuszczyłem. Dlaczego użyłem żargonu? Ponieważ są to terminy oznaczające wartościowe pojęcia, które dopiero niedawno stały się dostępne dla ludzi rozważających zagadnienie umysłu. Żadne inne słowa nie wyrażają tych pojęć rzeczowo, a zdecydowanie warto je poznać. Zatem, odwołując się najpierw do krótkiej historycznej dygresji, przedstawię je i umieszczę w kontekście, w którym będziemy ich używać.

Dwaj z najważniejszych wynalazców komputera to matematyk Alan Turing i matematyk i fizyk John von Neumann, Amerykanin węgierskiego pochodzenia. Choć Turing miał mnóstwo praktycznych doświadczeń z pierwszej ręki – projektował i konstruował wyspecjalizowane maszyny elektroniczne do łamania szyfrów, a maszyny te pomogły aliantom wygrać II wojnę światową – to erę komputerów rozpoczęła jego abstrakcyjna, czysto teoretyczna praca przedstawiająca jego koncepcję *uniwersalnej maszyny Turinga*. Von Neumann wykorzystał abstrakcję Turinga (która była naprawdę „filozoficzna" – to eksperyment myślowy, nie propozycja inżynieryjna) i na tyle ukonkretnił, że można było ją zmienić w (nadal dosyć abstrakcyjny) projekt rzeczywistego, praktycznego komputera elektronicznego. Ten abstrakcyjny projekt, znany jako *architektura von Neumanna*, można znaleźć w niemal każdym współczesnym komputerze na świecie, od ogromnych komputerów centralnych do układów scalonych

znajdujących się w sercu większości skromnych komputerów domowych.

Komputer ma podstawową *stałą lub sprzętową* architekturę, lecz też dużą plastyczność dzięki *pamięci*, która może przechowywać zarówno *programy* (inaczej znane jako *oprogramowanie*), jak i *dane*, nietrwałe wzorce mające śledzić to, co jest do zaprezentowania. Komputery, tak jak mózgi, mają więc niekompletną konstrukcję w momencie narodzin, ale są na tyle elastyczne, że można w nich wytworzyć określoną architekturę, wyspecjalizowaną maszynę, a każda taka maszyna będzie miała zaskakująco indywidualny tryb przyjmowania stymulacji z otoczenia (za pośrednictwem klawiatury lub innych urządzeń wejściowych), w końcu dających odpowiedzi (na monitorze czy innych urządzeniach wyjściowych).

Te tymczasowe struktury są „zrobione z reguł, a nie z kabli", a informatycy nazywają je *maszynami wirtualnymi*[68]. Maszynę wirtualną uzyskuje się, gdy szczególny wzorzec reguł (bardziej dosłownie: dyspozycje lub prawidłowości przejść) ogranicza całą tę plastyczność. Weźmy przykład osoby, która złamała rękę i ma na niej gips. Gips znacznie ogranicza ruchy jej ręki, a jego waga i kształt również wymagają dostosowania od reszty ruchów ciała tej osoby. A teraz weźmy mima (na przykład Marcela Marceau) imitującego kogoś z gipsem na ręce; jeśli mim dobrze potrafi pokazać tę sztuczkę, ruchy jego ciała będą ograniczone w niemal ten sam sposób; ma na swojej ręce *wirtualny gips* – i będzie on „niemal niewidzialny". Każdy, kto umie używać edytora tekstu, zna co najmniej jedną maszynę wirtualną, a jeśli używa kilku różnych edytorów tekstu lub korzysta z arkusza kalkulacyjnego czy gra w jakąś grę na tym samym komputerze, na którym korzysta z edytora tekstów, to zna kilka maszyn wirtualnych, które na zmianę istnieją w konkretnej maszynie rzeczywistej. Różnice przedstawione są w sposób bardzo widoczny, aby użytkownik wiedział, z którą maszyną wirtualną ma do czynienia w każdym momencie.

Wszyscy wiemy, że różne programy dają komputerom różne możliwości, jednak nie każdy zna szczegóły. Kilka z nich jest ważnych dla naszej historii, więc proszę o wyrozumiałość, bo krótko i elementarnie przedstawię wynalazek Alana Turinga.

Turing nie starał się wynaleźć edytora tekstu ani gry wideo, gdy dokonywał swoich wspaniałych odkryć. Myślał, samoświadomie i introspekcyjnie, o tym, jak on, matematyk, rozwiązuje problemy matematyczne czy też przeprowadza *obliczenia*, a następnie rozłożył *ciągi aktów myślowych* na ich prostsze składniki. „Co takiego robię – musiał zapytać – gdy wykonuję obliczenia? Cóż, najpierw pytam siebie, którą regułę zastosować, a następnie stosuję tę regułę, potem zapisuję wynik, patrzę na niego, a wreszcie pytam siebie, co dalej..." Turing był wyjątkowo zorganizowanym myślicielem, lecz jego strumień świadomości, jak mój, twój czy Jamesa Joyce'a, był bez wątpienia pstrokatą mieszaniną obrazów, decyzji, przeczuć, przypomnień i tym podobnych, z których udało mu się wydestylować matematyczną esencję: elementarną, minimalną sekwencję operacji, które mogły zrealizować cele osiągane przez niego w barwnych i meandrujących czynnościach jego świadomego umysłu. Rezultatem było udokumentowanie tego, co teraz znamy jako „maszyna Turinga", czyli genialna idealizacja i uproszczenie hiperracjonalnego, hiperintelektualnego zjawiska: matematyka wykonującego rygorystyczne obliczenia. Podstawowa idea miała pięć elementów:

(1) proces szeregowy (zdarzenia następują kolejno) w

(2) ściśle ograniczonej *przestrzeni roboczej*, do której

(3) dostarcza się zarówno *dane*, jak i *instrukcje*

(4) z biernej, ale całkowicie niezawodnej *pamięci*,

(5) a następnie działa na nich skończona lista *operacji elementarnych*.

W oryginalnym sformułowaniu Turinga przestrzeń robocza była wyobrażona jako czytnik znajdujący się tylko nad jedną komórką papierowej taśmy i stwierdzający, czy jest w niej zapisane zero, czy jeden. W zależności od tego, co „zobaczył", wymazywał zero lub jeden i drukował inny symbol bądź pozostawiał komórkę bez zmian. Następnie przesuwał taśmę w lewo lub w prawo o jedną komórkę i spoglądał ponownie, w każdym przypadku kierując się skończoną listą instrukcji sprzętowych, które wchodziły w skład jego *tablicy maszyny*. Taśma była pamięcią.

Lista elementarnych operacji Turinga (czy, jak kto woli, aktów „nierozkładalnych dalej w introspekcji") została celowo zubożona, aby nie było wątpliwości co do ich mechanicznej realizowalności. Innymi słowy, dla matematycznych celów Turinga ważne było, aby nie pozostały niedopowiedzenia, aby każdy krok w procesie, który badał, był tak prosty, tak banalny, że mógłby zostać wykonany przez głupka – kogoś, kogo mogłaby zastąpić maszyna: CZYTAJ, USUŃ, DRUKUJ, PRZESUŃ W LEWO O JEDNO MIEJSCE i tak dalej.

Wiedział oczywiście, że jego idealna dokumentacja mogła posłużyć pośrednio jako plan faktycznej maszyny obliczeniowej, i zauważyli to inni, szczególnie von Neumann, który zmodyfikował podstawowe idee Turinga, tworząc abstrakcyjną architekturę pierwszego praktycznie wykonalnego komputera cyfrowego. Tę architekturę nazywamy *maszyną von Neumanna*.

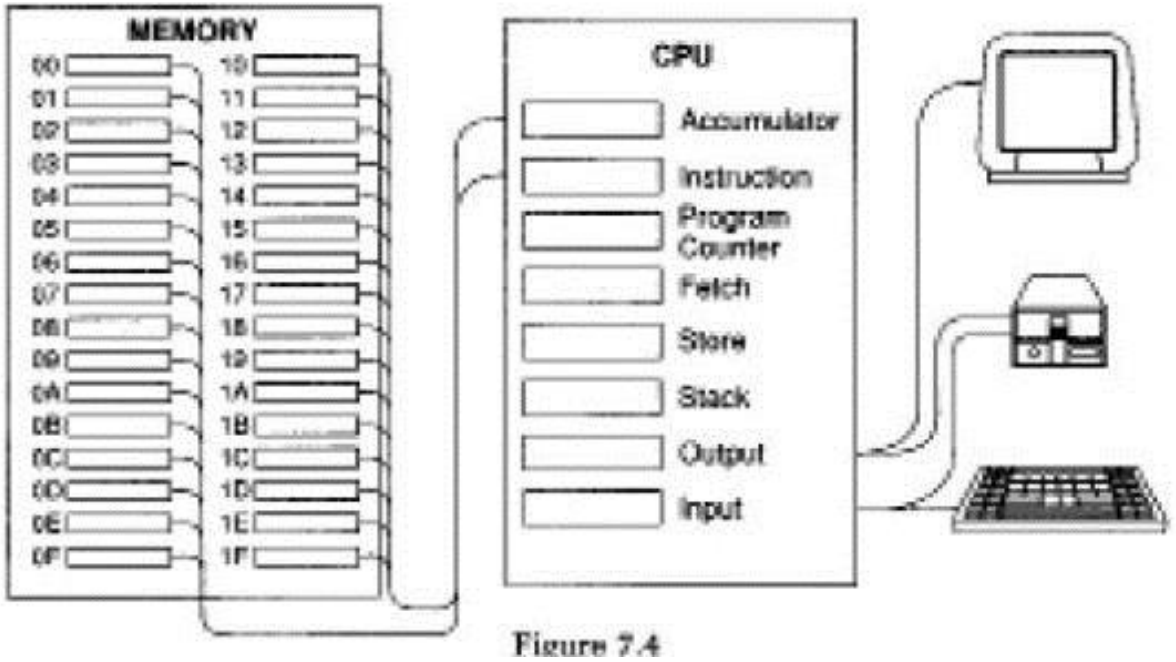


Figure 7.4

Ryc. 7.4

Z lewej strony widać *pamięć* czy też *pamięć o dostępie bezpośrednim* (RAM – *random access memory*), w której przechowywane są zarówno dane, jak i instrukcje, zakodowane jako ciąg cyfr binarnych lub „bitów", na przykład 00011011 lub 01001110. Szeregowy proces Turinga przebiega w przestrzeni roboczej składającej się z dwóch „rejestrów" oznaczonych jako

akumulator oraz *rejestr instrukcji*. Instrukcja zostaje skopiowana elektronicznie do rejestru instrukcji, który następnie ją *wykonuje*. Na przykład, jeśli instrukcja (przetłumaczona na język polski) mówi „wyczyść akumulator", komputer wstawia liczbę 0 do akumulatora, a jeśli instrukcja mówi „dodaj zawartość rejestru pamięci 07 do liczby w akumulatorze", komputer pobierze liczbę z rejestru pamięci o adresie 07 (zawartością może być *dowolna* liczba) i doda ją do liczby w akumulatorze. I tak dalej. Jakie są tu operacje elementarne? Zasadniczo operacje arytmetyczne: dodaj, odejmij, pomnóż i podziel; operacje przenoszenia danych: pobierz, zapisz, przekaż do wyjścia, pobierz z wejścia; oraz (serce „logiki" komputera) instrukcje *warunkowe*, jak na przykład „JEŚLI liczba w akumulatorze jest większa od zera, TO przejdź do instrukcji w rejestrze 29; w innym wypadku przejdź do następnej instrukcji". W zależności od modelu komputera może istnieć nawet tylko 16 operacji elementarnych lub ich setki, a wszystkie zapisane są w wyspecjalizowanych układach. Każda operacja elementarna jest zakodowana przez unikatowy wzorzec binarny (na przykład DODAJ może mieć kod 1011, a ODEJMIJ – 1101), a w momencie, w którym te konkretne ciągi wylądują w rejestrze instrukcji, stają się swego rodzaju wybieranymi numerami telefonów, które mechanicznie otwierają linie do odpowiedniego wyspecjalizowanego układu – układu dodawania, odejmowania itd. Oba rejestry, w których tylko jedna instrukcja i jedna wartość może pojawić się w danym momencie, są znanym „wąskim gardłem von Neumanna", miejscem, w którym wszystkie czynności systemu muszą przejść pojedynczo przez wąską lukę. Na szybkim komputerze mogą przebiegać miliony takich operacji na sekundę, a w milionach dają pozornie magiczne efekty, zauważalne przez użytkownika.

Wszystkie komputery cyfrowe są bezpośrednimi potomkami tego projektu; a mimo wprowadzenia wielu modyfikacji i ulepszeń, jak wszystkie kręgowce, łączy je fundamentalna, ukryta architektura. Podstawowe operacje, tak arytmetyczne z pozoru, nie wydają się z początku szczególnie pokrewne podstawowym „operacjom" zwykłego strumienia świadomości – myśleniu o Paryżu, zadowoleniu z zapachu chleba z pieca, zastanawianiu się, dokąd pojechać na wakacje – ale nie martwiło to Turinga czy von Neumanna. Ważne było dla nich to, że ten ciąg czynności mógł „zasadniczo" zostać rozbudowany, aby zawrzeć *wszelką* „racjonalną myśl", a być może również wszystkie „nieracjonalne myśli". Cóż za ironia losu, że ta architektura od czasów powstania jest błędnie opisywana przez prasę popularną. Te nowe, fascynujące maszyny von Neumanna zostały nazwane „ogromnymi mózgami elektronicznymi", a w rzeczywistości były *ogromnymi umysłami elektronicznymi*, elektronicznymi imitacjami – znacznymi uproszczeniami – tego, co William James nazwał strumieniem świadomości, meandrującym ciągiem świadomych treści umysłowych, wspaniale przedstawionym przez Jamesa Joyce'a w jego powieściach. Natomiast architektura mózgu jest zdecydowanie równoległa, z milionami jednocześnie aktywnych torów działania. Musimy zrozumieć, jak Joyce'owskie (lub, jak powiedziałem, von Neumannowskie) szeregowe zjawisko może zaistnieć, ze wszystkimi dobrze znanymi osobliwościami, w równoległym zgiełku mózgu.

Oto zły pomysł: nasi przodkowie, hominidy, musieli myśleć w sposób bardziej wysublimowany, logiczny, więc dobór naturalny stopniowo zaprojektował oraz zainstalował sprzętową maszynę von Neumanna w lewej („logicznej", „świadomej") półkuli ludzkiej kory mózgowej. Mam nadzieję, że powyższa narracja jasno wskazuje, że chociaż pomysł ten jest *logicznie* możliwy, to jest on zupełnie niewiarygodny biologicznie – nasi przodkowie równie dobrze mogli dostać skrzydeł lub urodzić się z pistoletami w dłoniach. Nie tak działa ewolucja.

Wiemy, że w mózgu jest coś przynajmniej *trochę* podobnego do maszyny von Neumanna, ponieważ „przez introspekcję" wiemy, że mamy świadome umysły, a umysły, które tym samym odkrywamy, są przynajmniej pod następującym względem podobne do maszyny von Neumanna: były dla niej inspiracją! Ten historyczny fakt pozostawił szczególnie przekonujący ślad:

programiści komputerowi powiedzą ci, że jest diabelsko trudno programować komputery równoległe, nad którymi się obecnie pracuje, a stosunkowo łatwo zaprogramować szeregową maszynę von Neumanna. Gdy programuje się konwencjonalną maszynę von Neumanna, wiadomo, na czym się oprzeć; gdy zaczyna się robić trudno, wystarczy spytać siebie: „Co bym zrobił, gdybym był maszyną próbującą rozwiązać ten problem?", a to prowadzi do odpowiedzi w rodzaju: „Cóż, najpierw zrobiłbym to, a następnie musiałbym zrobić tamto" itd. Gdy spytasz siebie: „Co zrobiłbym w tej sytuacji, gdybym był procesorem równoległym, mającym tysiąc torów przetwarzania?", nie masz pojęcia, co odpowiedzieć; nie znasz z autopsji – nie masz żadnego bezpośredniego dostępu do – procesu zachodzącego w tysiącu torów jednocześnie, mimo że jest to coś, co przebiega w twoim mózgu. Twój jedyny dostęp do tego, co się w nim dzieje, odbywa się w „formacie" sekwencyjnym, niesamowicie przypominającym architekturę von Neumanna – choć przedstawianie tego w ten sposób jest anachroniczne.

Jak już widzieliśmy, istnieje duża różnica między (standardową) szeregową architekturą komputera oraz równoległą architekturą mózgu. Ten fakt jest często cytowany jako zarzut wobec sztucznej inteligencji, w której próbuje się stworzyć inteligencję podobną do ludzkiej przez tworzenie programów (prawie zawsze) wykorzystujących maszynę von Neumanna. Czy różnica architektury jest różnicą teoretycznie istotną? W pewnym sensie nie. Turing udowodnił – i jest to prawdopodobnie jego największe osiągnięcie – że jego maszyna uniwersalna jest w stanie obliczyć każdą funkcję, którą obliczy jakikolwiek inny komputer o jakiejkolwiek architekturze. W rezultacie uniwersalna maszyna Turinga jest idealnym matematycznym kameleonem, który potrafi naśladować *każdą* inną maszynę obliczeniową i kiedy to robi, wykonać *dokładnie to*, czego tamta maszyna dokonuje. Wystarczy podać uniwersalnej maszynie Turinga odpowiedni opis innej maszyny i – jak Marcel Marceau (uniwersalna maszyna pantomimiczna) wyposażony w opisaną dobrze choreografię – niezwłocznie naśladuje ona działanie zgodnie z tym opisem; praktycznie staje się tą maszyną. Program komputerowy może być więc postrzegany jako lista instrukcji elementarnych do wykonania lub jako opis maszyny do naśladowania.

Czy potrafisz naśladować Marcela Marceau naśladującego pijaka naśladującego baseballistę? Możesz stwierdzić, że najtrudniejsze w tej sztuczce jest śledzenie różnych poziomów naśladowania, ale maszynom von Neumanna przychodzi to naturalnie. Gdy masz już maszynę, na której możesz się oprzeć, możesz umieścić w niej wirtualne maszyny jak matrioszki. Na przykład początkowo możesz zmienić swoją maszynę von Neumanna w maszynę uniksową (system operacyjny Unix), a następnie zaimplementować maszynę Lispa (język programowania Lisp) na maszyny uniksowe – razem z programem WordStar, Lotus 123 i wieloma innymi maszynami wirtualnymi – a następnie zaimplementować na maszyny Lispa komputer grający w szachy. Każda maszyna wirtualna jest rozpoznawana poprzez swój *interfejs użytkownika* – jej wygląd na ekranie monitora i reakcje na dane wejściowe – a ta autoprezentacja często nazywana jest *iluzją użytkownika*, gdyż użytkownik nie może stwierdzić – i nie obchodzi go – w jaki sposób konkretna maszyna wirtualna, z której korzysta, jest zaimplementowana w sprzęcie. Nie jest dla niego istotne, czy maszyna wirtualna znajduje się jeden, dwa, trzy czy dziesięć poziomów od maszyny sprzętowej[69]. (Na przykład użytkownicy programu WordStar mogą go rozpoznać i wchodzić w interakcję z wirtualną maszyną WordStar, gdziekolwiek ją znajdą, bez względu na to, jakiego rodzaju sprzęt ją implementuje).

Zatem maszyna wirtualna jest tymczasowym zbiorem prawidłowości o złożonej strukturze, rządzących dodatkowo podstawowym sprzętem za pośrednictwem *programu*: przepisu o złożonej strukturze, składającego się z setek tysięcy instrukcji, które dają temu sprzętowi ogromny, spójny zbiór nawyków czy dyspozycji do reagowania. Jeśli spojrzysz na szczegóły wszystkich tych instrukcji przewijających się przez rejestr instrukcji, nie obejmujesz

pełnego obrazu sytuacji; jednak gdy zrobisz krok w tył, funkcjonalna architektura wyłaniająca się z wszystkich tych mikroustawień będzie wyraźnie widoczna: składa się z wirtualnych *rzeczy*

takich jak fragmenty tekstu, kursory, gumki, spraye do malowania, dokumenty

oraz z wirtualnych *miejsc*

takich jak katalogi, menu, ekrany, powłoki systemowe

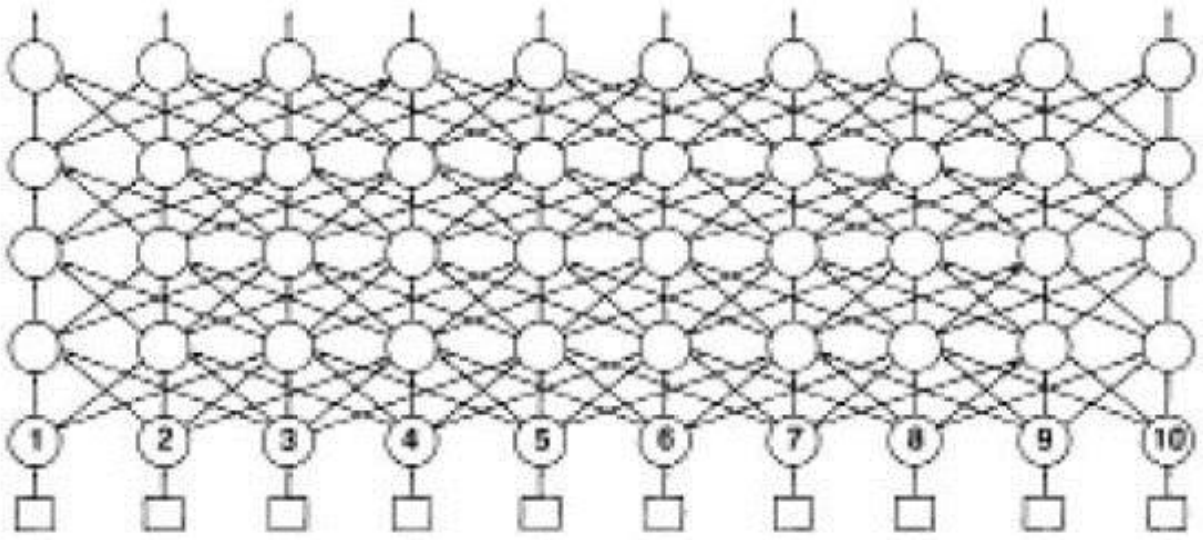
połączonych ze sobą wirtualnymi *ścieżkami*

takimi jak „klawisz [Esc] powoduje wyjście do systemu DOS" lub *wejście* do menu DRUKOWANIA z menu GŁÓWNEGO

i pozwalających na wykonanie istotnych i interesujących *operacji wirtualnych*

takich jak szukanie słowa w dokumencie lub powiększenie prostokąta narysowanego na ekranie.

Stąd, iż każda maszyna obliczeniowa może być naśladowana przez maszynę wirtualną na maszynie von Neumanna, wynika, że jeśli mózg jest potężną maszyną o przetwarzaniu równoległym, to również może zostać idealnie naśladowany przez maszynę von Neumanna. I od samego zarania ery komputerowej teoretycy używali też kameleonowej mocy maszyny von Neumanna, aby stworzyć *wirtualne* architektury równoległe, które miały utworzyć struktury podobne do mózgowych[70]. Jak sprawić, aby maszyna wykonująca jedną czynność naraz stała się maszyną wykonującą wiele czynności jednocześnie? Podobnie jak robi się na drutach. Załóżmy, że procesor równoległy, który symulujemy, ma dziesięć torów. Najpierw maszyna von Neumanna dostaje instrukcje, aby wykonać operacje obsługiwane przez pierwszy węzeł pierwszego toru (węzeł 1 na rycinie 7.5) i zapisać rezultat w pamięci „buforowej", a następnie węzeł 2 i tak dalej, aż wszystkie dziesięć węzłów w pierwszej warstwie posuwa się jeden krok naprzód. Następnie maszyna von Neumanna zajmuje się wynikami każdego z tych rezultatów pierwszej warstwy w kolejnej warstwie węzłów, pobierając wcześniej obliczone rezultaty po kolei z pamięci buforowej i stosując je jako dane wejściowe w kolejnej warstwie. Pracowicie posuwając się do przodu, robi na drutach, wymienia czas na przestrzeń. Maszyna wirtualna o dziesięciu torach będzie potrzebowała *przynajmniej* dziesięć razy więcej czasu na symulację niż maszyna jednotorowa, a maszyna o milionie torów (taka jak mózg) będzie wymagała przynajmniej milion razy więcej czasu na symulację. Dowód Turinga nie mówi nic o prędkości symulacji, a w przypadku architektur pewnego rodzaju nawet współczesne komputery cyfrowe o zawrotnej prędkości nie poradzą sobie z takim zadaniem. Właśnie dlatego naukowcy zajmujący się sztuczną inteligencją, zainteresowani rozwijaniem mocy architektury równoległej, zaczynają dziś interesować się *prawdziwymi* maszynami równoległymi – tworami, które można by zasadniej nazywać „gigantycznymi mózgami elektronicznymi" – na których tworzą swoje symulacje. Jednak zasadniczo każda maszyna równoległa może zostać idealnie – choć niewydajnie – symulowana jako maszyna wirtualna na szeregowej maszynie von Neumanna[71].



Ryc. 7.5

Jesteśmy teraz gotowi, aby wywrócić tę standardową ideę do góry nogami. Tak jak można symulować równoległy mózg na szeregowej maszynie von Neumanna, tak też można, zasadniczo, symulować maszynę von Neumanna (lub coś w jej rodzaju) na równoległym sprzęcie, i właśnie to sugeruję: świadome ludzkie umysły są mniej więcej szeregowymi maszynami zaimplementowanymi – niewydajnie – na równoległym sprzęcie, który zapewniła nam ewolucja.

Co liczy się jako „program", gdy mówimy o wirtualnej maszynie działającej na równoległym sprzęcie mózgowym? Ważne jest istnienie dającej się wykorzystać plastyczności, która może przyjąć niezliczoną ilość różnych mikronawyków, a tym samym makronawyków. W przypadku maszyny von Neumanna jest to osiągane przez setki tysięcy zer i jedynek (bitów), podzielonych na „słowa" 8-, 16-, 32- lub 64-bitowe, w zależności od maszyny. Słowa są przechowywane osobno w rejestrach pamięci, a rejestr instrukcji może uzyskać jednorazowo dostęp do jednego słowa. W przypadku maszyny równoległej możemy przypuszczać, że jest to osiągane przez tysiące, miliony czy miliardy ustawień siły połączeń między neuronami, które łącznie zapewniają bazowemu sprzętowi nowy zestaw makronawyków, nowy zestaw warunkowych prawidłowości zachowania.

Jak zatem te programy złożone z milionów ustawień siły połączeń neuronowych zostają zainstalowane na komputerze mózgu? W maszynie von Neumanna po prostu „ładuje się" program z dysku do pamięci głównej i stąd komputer dostaje natychmiastowy zestaw nowych nawyków; w przypadku mózgów potrzeba *uczenia*, w tym szczególnie powtarzalnej autostymulacji podobnej do przedstawionej w podrozdziale 5. Oczywiście, analogia nie jest tutaj pełna. Procesor sztywno odpowiada na ciąg bitów, które tworzą jego słowa, traktując je jak instrukcje w jego całkowicie własnym i ustalonym *języku maszynowym*. Jest to charakterystyczne dla *programowalnego komputera cyfrowego*, a ludzki mózg nie jest niczym takim. Jest prawdopodobnie prawdą, że każde ustawienie siły połączeń między neuronami w mózgu ma określony wpływ na wynikowe zachowanie otaczającej sieci, lecz nie ma żadnego powodu, aby sądzić, że dwa różne mózgi będą miały „ten sam system" połączeń wewnętrznych, a więc prawie na pewno nie ma tu nic analogicznego do stałego języka maszynowego, który, powiedzmy, mają wszystkie komputery IBM i zgodne z IBM. Jeśli zatem co najmniej dwa mózgi „mają wspólne

oprogramowanie", to nie na mocy prostego, bezpośredniego procesu analogicznego do kopiowania programu w języku maszynowym z jednej pamięci do drugiej. (Co więcej, oczywiście plastyczność odpowiedzialna za pamięć w mózgu nie jest wydzielona jako bierny magazyn; podział pracy między pamięcią a procesorem jest artefaktem bez analogii w mózgu, a do tematu tego wrócimy w rozdziale 9).

Skoro istnieją tak istotne – a często pomijane – różnice, dlaczego obstaję przy porównaniu ludzkiej świadomości do oprogramowania? Ponieważ, jak mam nadzieję pokazać, pewne ważne, a inaczej nadzwyczaj zagadkowe cechy świadomości mogą zostać dobrze wyjaśnione na mocy hipotezy, że ludzka świadomość (1) jest innowacją zbyt świeżą, aby była wbudowana we wrodzoną maszynerię, (2) jest w dużej mierze wytworem ewolucji kulturowej i zostaje przekazana mózgom we wczesnym uczeniu, oraz (3) jej pomyślna instalacja zależy od niezliczonych mikroustawień plastycznego mózgu, co oznacza, że jest bardzo prawdopodobne, iż jej funkcjonalnie istotne cechy są niewidoczne w badaniach neuroanatomicznych pomimo dostrzegalności jej efektów. Tak jak żaden informatyk nie próbowałby zrozumieć różnych silnych i słabych stron programów WordStar i WordPerfect przez wnioskowanie na podstawie informacji o różnicach we wzorcach napięcia w pamięci, tak też żaden kognitywista nie powinien oczekiwać, że ludzką świadomość można zrozumieć po prostu na bazie neuroanatomii. Poza tym (4) idea *iluzji użytkownika* maszyny wirtualnej jest kusząco sugestywna: jeśli świadomość jest maszyną wirtualną, kto jest użytkownikiem, dla którego działa iluzja użytkownika? Przyznaję, że wygląda to podejrzanie, jakbyśmy nieubłaganie zmierzali z powrotem do jaźni kartezjańskiej, siedząc przy korowej stacji roboczej i reagując na iluzję użytkownika oprogramowania w niej działającego, ale jak zobaczymy, są pewne sposoby ucieczki od tego okropnego rozwiązania.

Załóżmy więc na chwilę, że istnieje lepiej lub gorzej zaprojektowana (i pozbawiona błędów) wersja tej maszyny wirtualnej strumienia świadomości – maszyny Joyce'owskiej – w memosferze. Jak widzieliśmy, skoro nie ma wspólnego języka maszynowego mózgów, metody przekazu gwarantujące w miarę jednorodną maszynę wirtualną działającą poprzez kulturę muszą być społeczne, wysoce wrażliwe na kontekst i do pewnego stopnia samoorganizujące oraz samokorygujące. „Porozumienie" dwóch różnych komputerów – na przykład Macintosh i PC IBM – wymaga skomplikowanej i drobiazgowej inżynierii, opartej na precyzyjnych informacjach na temat wewnętrznej maszynerii obu tych systemów. O ile ludzie mogą „dzielić się oprogramowaniem" bez posiadania jakiejkolwiek wiedzy tego typu, musi się tak dziać dlatego, że wspólne systemy mają wysoki stopień elastyczności i tolerancji na format. Istnieje kilka metod dzielenia się takim oprogramowaniem: uczenie się przez naśladownictwo, uczenie się w wyniku wzmacniania (albo celowo stosowanego przez nauczyciela – nagroda, zachęta, dezaprobata, groźba, albo subtelnie i nieświadomie przekazanego w ramach komunikacji) oraz uczenie się jako rezultat umyślnego nauczania w języku naturalnym, który już został opanowany poprzez dwie pierwsze metody. (Przykładem pierwszych metod są tego rodzaju nawyki, które zostałyby wszczepione przez częste powtarzanie do nowicjusza: „Powiedz mi, co teraz robisz" oraz „Powiedz mi, dlaczego to robisz". Stąd potem nawyk kierowania tych samych próśb do siebie).

Przypuszczam, że nie tylko język mówiony, ale i pismo odgrywa ważną rolę w rozwoju i dopracowaniu maszyn wirtualnych, które u większości z nas działają prawie cały czas w mózgach. Tak jak koło jest doskonałym wynalazkiem techniki, a jego funkcjonowanie zależy od torów, utwardzonych dróg czy innych sztucznie wyrównanych powierzchni, tak i maszyna wirtualna, o której mówię, może istnieć tylko w środowisku, w którym zachodzą nie tylko językowe i społeczne interakcje, ale również pismo i planowanie, po prostu dlatego, że ze względu na ich wymagania wobec pamięci i rozpoznawania wzorców ich implementacja wymaga „odciążenia" mózgu przez wykorzystanie buforów w otoczeniu jako pamięci. (Zauważmy, że

zakłada to, iż „umysł przedpiśmienny" może równie dobrze wykorzystać znacząco architektury wirtualne innego rodzaju niż te spotykane w społecznościach piśmiennych).

Wyobraźmy sobie dodawanie w głowie dwóch liczb dziesięciocyfrowych bez użycia papieru i ołówka *lub* mówienia na głos. Wyobraźmy sobie wymyślanie, bez obrazków, połączenia trzech autostrad w węźle drogowym typu koniczyna, tak aby osoba jadąca z któregokolwiek kierunku na którejkolwiek autostradzie w którymkolwiek kierunku na którąkolwiek autostradę nie miała obowiązku wjazdu na trzecią autostradę. Są to typy problemów, które ludzie bez trudu rozwiązują za pomocą zewnętrznych środków pamięci oraz przy użyciu istniejących wcześniej czytników (zwanych *oczami* i *uszami*), mających wysoce rozwinięte, wbudowane obwody rozpoznawania schematów. (Ciekawe obserwacje na ten temat można znaleźć w Rumelhart, rozdz. 14, w: McClelland i Rumelhart 1986).

Instalujemy w naszych mózgach zorganizowany i częściowo przetestowany zbiór nawyków umysłu, jak nazywa je politolog Howard Margolis (1987), w fazie wczesnego rozwoju w dzieciństwie. Możliwym szczegółom tej architektury przyjrzymy się bliżej w rozdziale 9, ale teraz wskazuję na to, że ogólna struktura nowego zestawu prawidłowości ma formę *szeregowego łańcucha*, w którym pierwsza „rzecz", a następnie kolejna „rzecz" odbywają się (z grubsza) w tym samym „miejscu". Ten strumień zdarzeń jest stworzony przez wiele wyuczonych nawyków, których najlepszym przykładem jest mówienie do siebie.

Skoro ta nowa maszyna, którą w sobie stworzyliśmy, jest kompleksem memów o wysokim stopniu replikacji, możemy zapytać, czemu zawdzięcza swój replikacyjny sukces. Powinniśmy oczywiście pamiętać, że *być może* nie jest dobra dla niczego – oprócz replikowania. *Może* być programem-wirusem, który szybko zaczyna pasożytować na ludzkich mózgach, nie dając istotom ludzkim, których mózgi zaraża, żadnej przewagi nad konkurencją. Co bardziej prawdopodobne, *pewne cechy* maszyny mogą być pasożytami istniejącymi tylko dlatego, że mogą, oraz dlatego, że nie jest możliwe – lub warte zawracania sobie głowy – pozbycie się ich. William James uważał, że byłoby niedorzecznością zakładać, iż najbardziej niesłychana rzecz, którą znamy we wszechświecie – świadomość – jest zwykłym artefaktem, nieodgrywającym żadnej istotnej roli w działalności naszego mózgu, ale choćby brzmiało to mało prawdopodobnie, nie można tego całkowicie wykluczyć, więc nie jest to tak naprawdę niedorzeczne. Istnieje mnóstwo dowodów świadczących o zyskach, które świadomość najwyraźniej nam daje, bez wątpienia zatem możemy podać różne racje jej bytu, jesteśmy jednak skłonni źle odczytywać te świadectwa, jeśli uważamy, że tajemnica pozostanie nieodkryta, jeżeli każda cecha świadomości nie ma – lub nie miała – funkcji (z *naszego* punktu widzenia jako „użytkowników" świadomości; Harnad 1982). Pozostaje miejsce na pewne surowe fakty bez jakiegokolwiek usprawiedliwienia funkcjonalnego. Pewne cechy świadomości mogą być po prostu egoistycznymi memami.

A jeśli mowa o korzyściach, to do rozwiązywania jakich problemów ta nowa maszyna najwyraźniej służy? Psycholog Julian Jaynes (1976) przekonująco argumentuje, że jej zdolności do samonapominania i samoprzypominania są warunkiem rodzaju rozwiniętych i długotrwałych procesów autokontroli, bez których nie mogłoby zostać zorganizowane rolnictwo, projekty budowlane oraz inne czynności cywilizowane i cywilizujące. Wydaje się, że również nieźle służy do swego rodzaju automonitorowania, które może uchronić uszkodzony system przed własnymi awariami; to wątek w dziedzinie sztucznej inteligencji, o którym pisał Douglas Hofstadter (1985). Psycholog Nicholas Humphrey (1976, 1983a, 1986) uważa, że świadomość służy do wykorzystywania czegoś, co mogłoby być nazwane „symulacjami społecznymi" – do korzystania z introspekcji w celu odgadywania, co inni myślą i czują.

U podstawy tych bardziej zaawansowanych i wyspecjalizowanych talentów leży zasadnicza zdolność do rozwiązywania metaproblemu, o czym następnie myśleć. Widzieliśmy na

początku tego rozdziału, że gdy organizm mierzy się z kryzysem (lub jedynie z trudnym i nowym problemem), może mieć własne środki, które byłyby bardzo przydatne w tych warunkach, *gdyby tylko mógł je odnaleźć i użyć ich na czas*! Reakcje orientujące, jak przypuszczał Odmar Neumann, mają tę zaletę, że mniej więcej wszyscy włączają się naraz, jednak to globalne pobudzenie, jak mogliśmy zaobserwować, jest tyleż problemem, co i rozwiązaniem. Nie pomaga ani trochę, chyba że w kolejnym kroku w mózgu pojawi się spójna aktywność od tych ochotników. Problem, którego częściowym rozwiązaniem były reakcje orientujące, polegał na uzyskaniu całkowitego, globalnego dostępu dla różnych grup specjalistów przyzwyczajonych do zajmowania się swoimi sprawami. Nawet jeśli dzięki bazowej architekturze w stylu Pandemonium chaos wkrótce się uspokaja, pozostawiając tymczasowo odpowiedzialnego jednego specjalistę (który być może, lepiej poinformowany przez konkurencję, wygrał), istnieje przynajmniej tyle złych rozwiązań tych konfliktów, ile jest dobrych. Nic nie gwarantuje, że politycznie najbardziej wydajny specjalista będzie odpowiedni na to stanowisko.

Platon dojrzał ten problem wyraźnie dwa tysiące lat temu i opisał go wspaniałą metaforą:

Więc zobacz, czy i wiedzy można też tak samo: nie mieć, chociaż się ją posiada – tak jak by kto dzikie ptaki – gołębie, czy jakieś inne – złowił, urządziłby im w domu gołębnik i tam by je chował. Więc w jakimś sposobie powiedzielibyśmy, że on je ma zawsze, bo je przecież posiada. Czy nie? [...] A w innym sposobie nie ma żadnego gołębia, tylko ma w stosunku do nich moc, skoro je w swoim ogrodzeniu pod ręką trzyma, żeby je, kiedy zechce, brać do ręki i mieć, i znowu je puszczać. To mu wolno robić, ile razy mu się podoba. [...] [T]ak teraz znowu zróbmy w każdej duszy pewnego rodzaju gołębnik; w nim różnorodne ptaki. Jedne stadami latają osobno od innych, inne po kilka sztuk razem, a inne pojedynczo, pomiędzy wszystkimi latają, którędy bądź. [*Teajtet*, s. 197, przeł. Władysław Witwicki]

Platon widział, że samo posiadanie ptaków nie wystarczy; trudną kwestią jest nauczenie się, w jaki sposób sprawić, by odpowiedni ptak przylatywał do ciebie, gdy go zawołasz. Twierdził też, że przez *rozumowanie* udoskonalamy umiejętność przywołania odpowiednich ptaków w odpowiednim czasie. Nauczenie się rozumowania jest więc opanowaniem technik przywoływania wiedzy[72]. I tu pojawiają się nawyki umysłu. Widzieliśmy już surowy zarys tego, jak takie ogólne nawyki umysłu, jak mówienie do siebie czy rysowanie dla siebie, *mogą* zwabiać odpowiednie kęsy informacji na powierzchnię (powierzchnię czego? – to temat, który poruszę w rozdziale 10). Jednak bardziej specyficzne nawyki umysłu, udoskonalenia i rozwinięcia konkretnych sposobów mówienia do siebie mogą jeszcze bardziej poprawić twoje szanse.

Filozof Gilbert Ryle w swojej pośmiertnie opublikowanej książce *On Thinking* (1979) stwierdził, że myślenie w rodzaju powolnego, trudnego rozmyślania, które najwyraźniej przedstawia pomnik *Myśliciel* autorstwa Auguste'a Rodina, musi w rzeczywistości być kwestią mówienia do siebie. Niespodzianka! Czy nie jest oczywiste, że to właśnie robimy, gdy myślimy? Cóż, tak i nie. Oczywiste jest, że jest to to, co (zwykle) wydaje się, że robimy; często możemy nawet powiedzieć innym różne słowa, które wyrażamy w naszych cichych monologach. Jednak wcale oczywiste nie jest, dlaczego mówienie do siebie jest w ogóle czymś dobrym.

Co robi *Le Penseur* w swoim, zdawałoby się, kartezjańskim wnętrzu? Lub, by wyrazić się bardziej naukowo, jak wyglądają procesy umysłowe, które odbywają się w kartezjańskiej *camera obscura*? [...] Powszechnie wiadomo, że niektóre nasze rozmyślania, ale nie wszystkie, kończą się rozwiązaniem naszych problemów; byliśmy we mgle, ale w końcu dostrzegliśmy rzeczy wyraźnie. Jednak jeśli czasem się udaje, to dlaczego nie zawsze? Jeśli poniewczasie, to dlaczego nie niezwłocznie? Jeśli z trudnością, to dlaczego nie łatwo? Dlaczego to w ogóle czasami działa? Jak to jest możliwe, że działa? [Ryle 1979, s. 65]

Nawyki umysłu były projektowane przez niezliczone wieki, *kształtując* przejścia na

wydeptanych ścieżkach eksploracji. Jak pisze Margolis:

[...] nawet istota ludzka dzisiaj (toteż tym bardziej daleki przodek współczesnych istot ludzkich) nie potrafi z łatwością i zwyczajnie utrzymać nieprzerwanej uwagi na jednym problemie dłużej niż przez kilka dziesiątek sekundy. A jednak zajmujemy się problemami, które wymagają zdecydowanie więcej czasu. Nasza metoda (jak możemy zaobserwować, przyglądając się sobie) wymaga okresów przemyśleń, po których następują okresy rekapitulacji, podczas których wyjaśniamy sobie to, co najwyraźniej wydarzyło się podczas przemyśliwania, co prowadzi do jakichś pośrednich rezultatów. Ma to oczywistą funkcję: przez powtarzanie sobie tych tymczasowych rezultatów [...] powierzamy je pamięci, gdyż bezpośrednie treści strumienia świadomości zostają szybko zagubione, jeśli ich nie powtórzymy. [...] Mając język, możemy samym sobie opisać, co takiego musiało się wydarzyć podczas rozmyślania, co doprowadziło do pewnego osądu, możemy stworzyć powtarzalną wersję procesu docierania do osądu i przekazać to do pamięci długotrwałej, tak naprawdę dzięki powtarzaniu. [Margolis 1987, s. 60]

To tutaj, w indywidualnych nawykach autostymulacji, powinniśmy szukać *prowizorki*. „Prowizorka" to termin używany przez hakerów komputerowych na oznaczenie prowizorycznego rozwiązania zwykle w programach w procesie debugowania, aby wszystko rzeczywiście działało. (Językoznawczyni Barbara Partee skrytykowała kiedyś nieelegancką łatę w programie sztucznej inteligencji zajmującym się analizą składniową zdań, gdyż była „dziwacznym rozwiązaniem hakera". Matka Natura pełna jest dziwacznych tworów rozwiązań hakerskich i powinniśmy się spodziewać ich znalezienia również w specyficznych sposobach użytkowania maszyny wirtualnej przez daną jednostkę).

Oto przekonujący przykład: ludzka pamięć nie jest w sposób wrodzony dobrze zaprojektowana jako absolutnie niezawodna pamięć operacyjna o dostępie bezpośrednim (czego potrzebuje każda maszyna von Neumanna), więc kiedy (kulturowo i czasowo rozproszeni) projektanci neumannowskiej wirtualnej maszyny stanęli przed zadaniem stworzenia stosownego substytutu w mózgu, napotkali różne sztuczki poprawiające pamięć. Podstawowe sztuczki to powtarzanie, powtarzanie i jeszcze raz powtarzanie, wspierane przez rymy i rytmiczne, łatwe do zapamiętania maksymy. (Rymy i rytmy wykorzystują ogromną moc istniejącego wcześniej systemu analizy słuchowej, który rozpoznaje wzorce w dźwiękach). Celowo powtarzane zestawienie elementów, między którymi trzeba było stworzyć połączenie – tak aby jeden z nich zawsze „przypominał" mózgowi o drugim – zostało ulepszone, jak możemy przypuszczać, jeszcze bardziej poprzez tworzenie maksymalnie bogatych skojarzeń, mających nie tylko cechy wizualne i słuchowe, ale wykorzystujących całe ciało. Marszczenie brwi i podpieranie podbródka Myśliciela, tak jak drapanie się w głowę, mamrotanie, chodzenie czy bazgroły, czy co tam indywidualnie lubimy, mogły się okazać nie tylko przypadkowymi skutkami ubocznymi świadomego myślenia, ale także funkcjonalnymi czynnikami (czy też śladowymi pozostałościami wcześniejszych, prostszych funkcjonalnych czynników) pracowitego dyscyplinowania mózgu, zamieniającego go w dojrzały umysł.

A zamiast dokładnych, systematycznych „cykli pobierania" czy „cykli instrukcji", które wprowadzają każdą nową instrukcję do ich rejestru, aby została wykonana, powinniśmy szukać niedoskonale wprowadzonych, trochę błądzących „reguł" przejścia, dalekich od logiki, gdzie głęboko wpisana w mózg skłonność do „swobodnego kojarzenia" powstaje na mocy długaśnych łańcuchów skojarzeń, które z grubsza zapewniają wypróbowanie odpowiednich ciągów. (W rozdziale 9 zajmiemy się rozwinięciem tej idei w sztucznej inteligencji; rozwinięcia z inaczej położonymi akcentami znajdziesz w Margolis 1987 oraz Calvin 1987, 1989. Zobacz również Dennett 1991b). Nie powinniśmy oczekiwać, że większość występujących ciągów okaże się sprawdzonymi *algorytmami*, gwarantującymi poszukiwane rezultaty, lecz jedynie częstszymi niż

przypadkowe udanymi polowaniami w ptaszarni Platona.

Analogia z maszynami wirtualnymi w informatyce stanowi użyteczny punkt widzenia na zjawisko ludzkiej świadomości. Komputery początkowo miały być jedynie maszynami matematycznymi, ale dziś ich zdolność obliczania jest wykorzystywana na tysiące pomysłowych sposobów i tworzy nowe maszyny wirtualne, takie jak gry wideo czy edytory tekstu, gdzie leżąca u podstaw umiejętność obliczania jest niemalże niewidoczna i gdzie nowe moce wydają się dosyć magiczne. Podobnie nasze mózgi nie zostały zaprojektowane (z wyjątkiem pewnych bardzo świeżych organów peryferyjnych) do redagowania tekstów, ale teraz duża część – być może największa – czynności, które odbywają się w dorosłym, ludzkim mózgu robi coś w rodzaju redagowania tekstów: realizacja i rozumienie mowy oraz szeregowe powtarzanie i reorganizowanie elementów językowych, czy lepiej mówiąc – ich neuronowych surogatów. A te czynności zwiększają i zmieniają leżące u ich podstaw moce sprzętowe na sposób, który wydaje się (z „zewnątrz") dosyć magiczny.

A jednak (jestem pewien, że chcecie się sprzeciwić): wszystko to ma niewiele albo i nic wspólnego ze świadomością! W końcu maszyna von Neumanna jest całkowicie nieświadoma; dlaczego zaimplementowanie jej – czy czegoś na jej kształt: maszyny Joyce'owskiej – miałoby być w ogóle świadome? Mam odpowiedź: maszyna von Neumanna, działając w ten sposób od samego początku, z maksymalnie wydajnymi łączami informacyjnymi, nie musiała stać się przedmiotem swoich własnych rozwiniętych systemów percepcyjnych. Natomiast działanie maszyny Joyce'owskiej jest dla niej samej tak samo „widzialne" i „słyszalne" jak każda inna rzecz w świecie zewnętrznym, do postrzegania której została zaprojektowana – z tego prostego powodu, że ta sama maszyneria percepcyjna koncentruje się na tych rzeczach.

Wydaje się to sztuczką z lustrami, wiem. I z pewnością jest nieintuicyjne, trudne do przełknięcia, z początku oburzające – dokładnie to, czego można by się spodziewać po pomyśle, który jest w stanie przebić się po wiekach tajemnic, sporów i zamętu. W kolejnych dwóch rozdziałach przyjrzymy się bliżej – i bardziej sceptycznie – jak ta pozorna sztuczka z lustrami może okazać się pełnoprawnym składnikiem wyjaśnienia świadomości.

Rozdział 8

Co robią z nami słowa?

Podobnie jak świadomość, mowa powstaje dopiero z potrzeby, z konieczności komunikowania się z innymi ludźmi.
Karol Marks, 1846 [przeł. Kazimierz Błeszyński]

Świadomość rozwinęła się w ogóle tylko pod naciskiem potrzeby powiadamiania się.
Friedrich Nietzsche, 1882 [przeł. Leopold Staff]

Zanim nie przyszedł do mnie mój nauczyciel, nie wiedziałam, kim jestem. Żyłam w świecie, który był nieświatem. Nie mogę liczyć, że uda mi się dobrze opisać ten nieświadomy, acz też świadomy okres nicości. [...] Nie miałam mocy myśli, więc nie porównywałam jednego stanu umysłu z innym.
Helen Keller, 1908

1. Rewizja: jedność uczyniona z wielości?

W rozdziale 5 naświetliliśmy uporczywie kuszący, zły pomysł teatru kartezjańskiego, w którym pokaz dźwięków i światła zostaje zaprezentowany samotnej, ale potężnej publice, Ego lub głównemu Kierownikowi. Mimo że sami dostrzegliśmy niespójność tej idei i zidentyfikowaliśmy jej konkurencję, model wielokrotnych szkiców, teatr kartezjański nadal będzie nas nawiedzał, dopóki trwale nie osadzimy naszego modelu na solidnej skale nauk empirycznych. Rozpoczęliśmy pracę nad tym zadaniem w rozdziale 6, a w rozdziale 7 zrobiliśmy kolejny krok do przodu. Powróciliśmy, dosłownie, do pierwszych zasad: zasad ewolucji, które były wskazówkami w spekulatywnej narracji na temat stopniowego procesu rozwoju konstrukcji, który wytworzył nasz rodzaj świadomości. Pozwoliło nam to spojrzeć na maszynerię świadomości od środka czarnej skrzynki – można by rzec: zza kulis – w hołdzie kuszącemu teatralnemu obrazowi, który próbujemy obalić.

W naszych mózgach znajduje się powiązany zbiór specjalistycznych obwodów mózgowych, które dzięki grupie nawyków wpojonych częściowo przez kulturę, a częściowo przez samoeksplorację współpracują ze sobą, tworząc mniej więcej uporządkowaną, mniej więcej efektywną, mniej więcej dobrze zaprojektowaną maszynę wirtualną, *maszynę joyce'owską*. Spajając te specjalistyczne organy, które wyewoluowały niezależnie od siebie, w imię jednego, wspólnego celu i tym samym nadając tej całości dalece potężniejsze moce, to oprogramowanie mózgu doprowadza do pewnego rodzaju wewnętrznego cudu politycznego: tworzy *wirtualnego kapitana* załogi bez nadawania żadnemu z nich długotrwałej władzy dyktatorskiej. Kto rządzi? Najpierw jedna koalicja, a potem kolejna, następując po sobie w sposób niechaotyczny dzięki dobrym metanawykom, które dostarczają spójnych, celowych ciągów, a nie niekończących się, bezładnych zamachów stanu.

Wynikająca z tego mądrość kierownicza jest po prostu jedną z mocy tradycyjnie przypisywanych Jaźni, ale jest ona istotna. William James uznał jej wagę, gdy wykpił ideę neuronu papieskiego w mózgu. Wiemy, że opis funkcji takiego szefującego podsystemu

w mózgu jest niespójny, wiemy jednak też, że obowiązki i decyzje kierownicze muszą być *jakoś* rozparcelowane w mózgu. *Nie jesteśmy* jak dryfujące statki z awanturniczymi załogami; nieźle radzimy sobie, nie tylko unikając mielizn i tym podobnych niebezpieczeństw, ale planując kampanie, poprawiając błędy taktyczne, rozpoznając subtelne zwiastuny możliwości oraz kontrolując ogromne projekty, trwające miesiące i lata. W kilku następnych rozdziałach bliżej przyjrzymy się architekturze maszyny wirtualnej, aby częściowo uzasadnić – nie ma tu ostatecznego dowodu – hipotezę, że rzeczywiście owa maszyna może realizować te kierownicze i inne funkcje. Zanim jednak to zrobimy, musimy ujawnić i zneutralizować kolejne źródło mistyfikacji: iluzję centralnego nadawacza sensu.

Jednym z głównych zadań wyimaginowanego szefa jest kierowanie komunikacją ze światem zewnętrznym. Jak widzieliśmy w rozdziale 4, idealizacja, która umożliwia heterofenomenologię, *zakłada*, że ktoś jest w środku i mówi, autor zapisu, nadawacz wszelkich sensów. Gdy zaczynamy interpretować gadatliwe głosowe dźwięki wydawane przez ciała, nie zakładamy, że są one przypadkowym jazgotem czy słowami wylosowanymi przez zakulisowych imprezowiczów, ale działaniami pojedynczego podmiotu działającego, (jednej jedynej) *osoby*, której ciało wytwarza te dźwięki. Jeśli w ogóle zdecydujemy się na interpretację, nie mamy innego wyjścia, jak tylko przyjąć istnienie osoby, której akty komunikacyjne interpretujemy. Nie jest to równoznaczne z przyjęciem istnienia *wewnętrznego* systemu, który jest szefem ciała, lalkarza pociągającego za sznurki, lecz jest to obraz narzucający się nam w sposób naturalny. Kuszące jest założenie, że ten wewnętrzny szef jest trochę jak prezydent Stanów Zjednoczonych, który może skierować rzecznika prasowego lub innych podwładnych do wystosowania komunikatów prasowych, gdy jednak te osoby się wypowiadają, to robią to w jego imieniu, wykonują jego akty mowy, za które on jest odpowiedzialny i których oficjalnie jest autorem.

W rzeczywistości nie istnieje taka hierarchia służbowa w mózgu, zarządzająca realizowaniem mowy (czy też pisma). W ramach rozmontowywania teatru kartezjańskiego trzeba podać bardziej realistyczne ujęcie faktycznego źródła (lub źródeł) aktów uznawania, pytania i innych aktów mowy, w naturalny sposób przypisywanych osobie, której ciało wypowiada się. Musimy zobaczyć, co się dzieje z niezbędnym w heterofenomenologii mitem, gdy uczyni się zadość zawiłościom realizacji języka.

Widzieliśmy już, jaki cień rzuca ten problem. W rozdziale 4 wyobrażaliśmy sobie robota Shakeya, który miałby elementarne zdolności do rozmowy, a przynajmniej do emitowania słów w pewnych warunkach. Założyliśmy, że Shakey mógł być zaprojektowany, aby „powiedzieć nam", jak odróżniał pudełka od piramid. Shakey mógłby powiedzieć: „Skanuję każdą sekwencję 10 000 zer i jedynek..." lub „Znajduję ciemno-jasne granice i rysuję w głowie białe linie..." czy też „Nie wiem; niektóre rzeczy po prostu wyglądają jak pudełka". Każde z tych „sprawozdań" było wysyłane z innego poziomu dostępu do wewnętrznych działań maszynerii rozpoznającej pudełka, który mogłaby mieć maszyneria wytwarzająca „sprawozdania", ale nie zagłębialiśmy się w to, w jaki sposób różne wewnętrzne stany maszyny wiązały się z rezultatami, do których doprowadzały. Był to celowo uproszczony model rzeczywistej realizacji języka, przydatny tylko do argumentacji na rzecz bardzo abstrakcyjnej tezy na podstawie tego eksperymentu myślowego: gdyby system emitujący zdania miał jedynie ograniczony dostęp do swoich stanów wewnętrznych oraz ograniczone słownictwo do tworzenia tych zdań, jego „sprawozdania" mogłyby być interpretowane jako prawdziwe tylko, jeśli odczytamy je metaforycznie. „Obrazy" Shakeya były przykładem tego, jak coś, co w rzeczywistości nie jest w ogóle obrazem, może być właśnie tym, o czym jednostka mówi, posługując się terminem „obraz".

Jedną kwestią jest otwarcie abstrakcyjnej możliwości; inną jest pokazanie, że ta możliwość urzeczywistnia się w naszym przypadku. To, co robił Shakey, nie było realnym

tworzeniem sprawozdań werbalnych, realnym mówieniem. Jak widzieliśmy, wyimaginowana werbalizacja Shakeya była swego rodzaju zwodniczym, „prefabrykowanym" językiem wbudowanym przez programatorów w oprogramowanie łatwe w obsłudze dla użytkownika. Chcesz sformatować dysk, a twój komputer „zadaje" ci przyjazne pytanie: „Czy na pewno chcesz to zrobić? Jeśli tak, z dysku zostanie usunięte wszystko! Wybierz T lub N". Bardzo naiwny musiałby być użytkownik, który sądziłby, że komputer naprawdę *chciał* być ostrożny.

Pozwólcie, że dam się wypowiedzieć krytykowi. Ten konkretny, wyimaginowany krytyk będzie wtrącał się w nasze dyskusje i badania w późniejszych rozdziałach, więc dam mu imię. Otto powiada tak:

Tanią sztuczką było odnoszenie się do Shakeya per „on" zamiast „to"; problem z Shakeyem jest taki, że nie ma prawdziwych wnętrzności tak jak my; nie ma nic, co jest do nich podobne. Nawet gdyby maszyneria, która pobierała informacje wejściowe z „oka" kamery telewizyjnej i zamieniła je w odróżnianie pudełek, była wyraźnie analogiczna do maszynerii w *naszym* systemie wzrokowym (a taka nie była) i nawet gdyby maszyneria kierująca realizacją ciągów angielskich słów była wyraźnie analogiczna do maszynerii w *naszych* systemach językowych, kierującej realizacją ciągów angielskich słów (a taka nie była), wówczas *nadal* czegoś by brakowało: pośrednika w każdym z nas, wyrażającego oceny, gdy mówimy o sobie samych. Problem z Shakeyem jest taki, że jego wejście i wyjście są ze sobą źle połączone – w sposób eliminujący obserwatora (osobę przeżywającą, osobę cieszącą się czymś), który musi znajdować się gdzieś między wejściem wzrokowym a wyjściem werbalnym, aby ktoś *miał na myśli* słowa Shakeya, gdy są one „wypowiadane".

Gdy *ja* mówię – kontynuuje Otto – mam na myśli to, co mówię. Moje świadome życie jest prywatne, ale mogę postanowić, że odkryję przed tobą pewne jego aspekty. Mogę zdecydować, że powiem ci o pewnych rzeczach dotyczących mojego obecnego lub wcześniejszego doświadczenia. Gdy to robię, formułuję zdania, które ostrożnie dostosowuję do materiału, z którego chcę zdać sprawę. Mogę poruszać się tam i z powrotem między przeżyciami a potencjalnym sprawozdaniem, konfrontując słowa z przeżyciem, upewniając się, że odnalazłem odpowiednią formę wyrażającą dokładnie to, co chcę. Czy wino ma w smaku aromat *grejpfruta*, czy może bardziej przypomina mi *jagody*? Czy byłoby lepiej powiedzieć, że wyższy ton brzmiał *głośniej*, czy po prostu wydawał on się *wyraźniejszy* lub *bardziej natężony*? Odnoszę się do mojego szczególnego, świadomego przeżycia i dokonuję osądu, które słowa najlepiej oddają jego charakter. Gdy jestem usatysfakcjonowany precyzyjnym wysłowieniem, wyrażam je. Z mojego introspekcyjnego raportu możesz dowiedzieć się o pewnej cesze mojego świadomego przeżycia.

Jako heterofenomenolodzy musimy podzielić ten tekst na dwie części. Z jednej strony mamy twierdzenia dotyczące tego, jakie wydaje się Ottonowi przeżycie mówienia. Są one niepodważalne; przeżycie wydaje się mu właśnie *takie* i musimy uznać je za informację wymagającą wyjaśnienia. Z drugiej strony mamy teoretyczne twierdzenia Ottona (czy są one wnioskami z milczących argumentów?) dotyczące tego, co to przeżycie mówienia mówi o tym, co się w nim dzieje – oraz jak to się różni od tego, co działo się na przykład w Shakeyu. Nie mają one szczególnego statusu, ale będziemy je traktować z szacunkiem należącym się wszelkim rozważnym twierdzeniom.

Mogę obstawać przy tym, że należy wyeliminować, a nie poszukiwać pośrednika, wewnętrznego obserwatora w teatrze kartezjańskim, ale nie mogę tak *po prostu* się go pozbyć. Jeśli nie ma centralnego nadawacza sensu, skąd bierze się sens? Musimy zastąpić go przekonującym ujęciem, mówiącym, jak wypowiedź, którą ktoś miał na myśli – *realny* raport, bez żadnego cudzysłowu – mogła zostać stworzona bez imprimatur samotnego centralnego nadawacza sensu. Jest to główny cel tego rozdziału.

2. Biurokracja kontra pandemonium

Jednym z trupów w szafie współczesnego językoznawstwa jest to, że poświęca ogromnie dużo uwagi słuchaniu, ale w dużej mierze pomija mówienie, o którym można by powiedzieć, że jest z grubsza połową języka, a także połową ważniejszą. Mimo że jest wiele szczegółowych teorii i modeli *percepcji* językowej oraz *rozumienia* usłyszanych wypowiedzi (ścieżki od fonologii przez składnię do semantyki i pragmatyki), to nikt – ani Noam Chomsky, ani żaden z jego rywali czy stronników – nie ma nic zdecydowanie istotnego (poprawnego bądź błędnego) do powiedzenia o systemach *realizacji* języka. To tak, jakby wszystkie teorie sztuki były teoriami *doceniania* sztuki i jakby nie było w nich słowa na temat artystów, którzy je stworzyli – jakby cała sztuka składała się z gotowych przedmiotów docenianych przez marszandów i kolekcjonerów.

Trudno stwierdzić, dlaczego jest właśnie tak. Wypowiedzi *są* gotowymi przedmiotami, od których zaczynamy proces. Jest jasne, czym jest surowy materiał czy wejście systemów percepcji i rozumienia: pewnego rodzaju falami w powietrzu bądź ciągami znaków na różnych powierzchniach płaskich. Co prawda, jest niejasne i sporne, co ma być ostatecznym wytworem procesu rozumienia, ale przynajmniej ten głęboki spór pojawia się na końcu badanego procesu, a nie na jego początku. Wyścig z jasną linią startową można przynajmniej racjonalnie rozpocząć, nawet jeśli nikt nie jest do końca pewien, gdzie się skończy. Czy „wyjście" lub „wytwór" rozumienia mowy to *dekodowanie* czy *tłumaczenie* wejścia na nowe reprezentacje – może na zdanie w języku „myśleńskim", a może na obrazek w głowie – czy jest to zbiór *głębokich struktur* lub jakiegoś dotychczas niewyobrażonego bytu? Językoznawcy mogą zdecydować o odłożeniu odpowiedzi na to trudne pytanie na później, podczas gdy zajmują się bardziej peryferyjnymi częściami tego procesu.

Jednak w przypadku realizacji mowy nikt jeszcze nie wypracował żadnego jasnego, powszechnie akceptowanego opisu tego, co rozpoczyna ten proces, prowadzący ostatecznie do pełnej wypowiedzi, trudne jest więc choćby rozpoczęcie rozważań nad jakąś teorią. Trudne, ale nie niemożliwe. Ostatnio pojawiły się dobre prace na temat tej kwestii, oparte na świetnie napisanej i usystematyzowanej przez holenderskiego psycholingwistę Pima Levelta książce *Speaking* (1989). Idąc wstecz od wytworu końcowego bądź też od środka w obu przeciwnych kierunkach, możemy dojrzeć sugestywne przebłyski maszynerii, która projektuje nasze wypowiedzi i sprawia, że zostają wyrażone. (Poniższe przykłady pochodzą z rozważań Levelta).

Mowa nie powstaje w „procesie wsadowym", który projektuje i realizuje jedno słowo naraz. Istnienie przynajmniej ograniczonej możliwości wybiegania w przyszłość w systemie ujawnia się w rozłożeniu akcentu w wypowiedzi. Prosty przypadek – akcent w słowie „sixteen" zależy od kontekstu:

ANDY: Ile dolarów to kosztuje?

BOB: Chyba szesnaście (ang. *sixTEEN*).

ANDY: Szesnaście (ang. *SIXteen*) dolarów to niedużo.

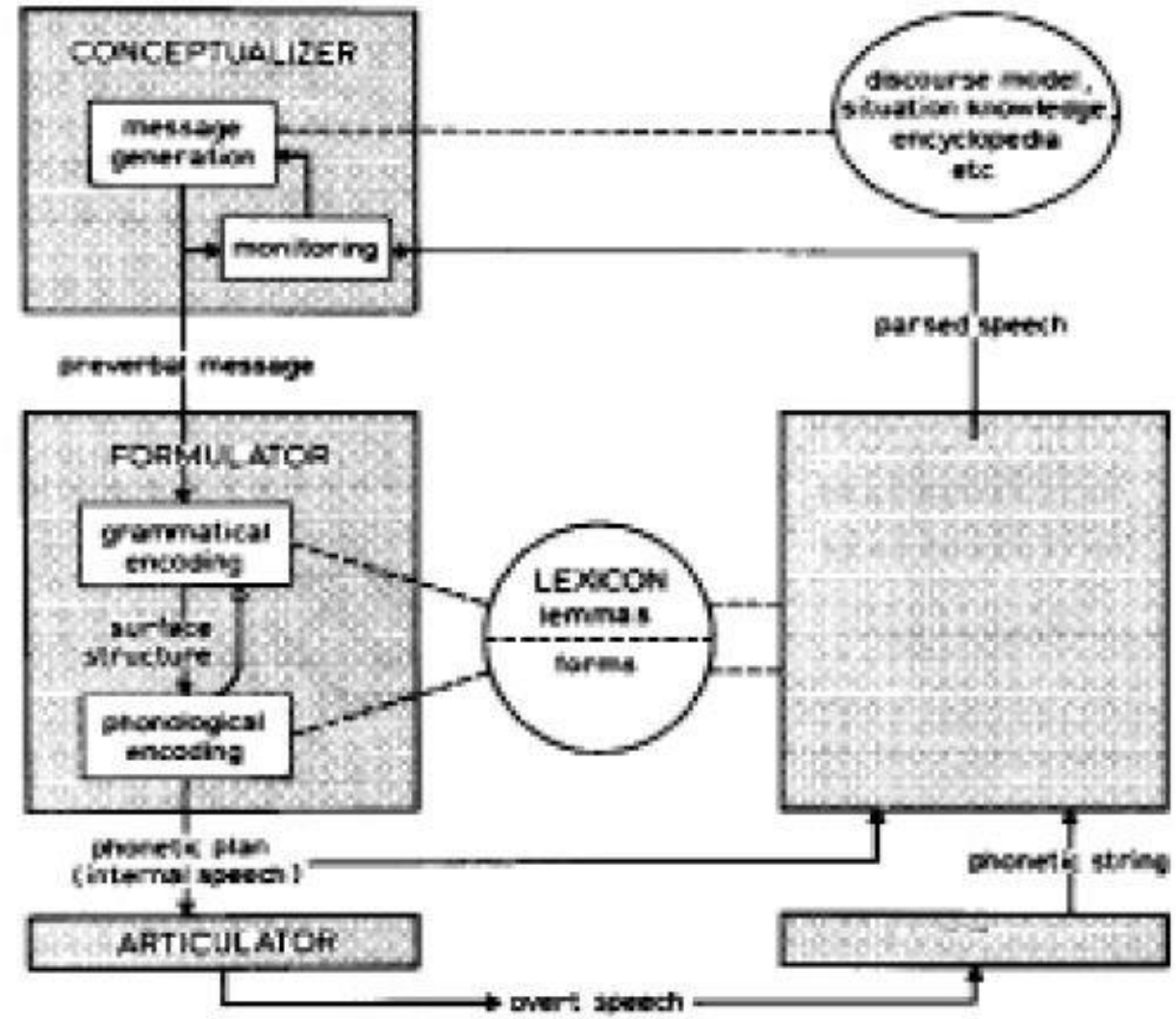
W przypadku drugiej wypowiedzi Andy'ego musi on dostosować wymowę „sixteen" do słowa, które następuje (ang. *DOLLars*). Gdyby chciał powiedzieć: „Szesnaście (ang. *sixTEEN*) to niedużo", inaczej rozłożyłby akcent w tym słowie. Kolejny przykład: zwróćmy uwagę na to, jak różni się akcentacja w następujących dwóch przypadkach w słowie „Tennessee": „Jechałem z Nashville w TennesSEE, do granicy TENnessee (ang. *TENnessee border*).

Spuneryzmy[73] i inne błędy mowy pokazują w sposób raczej rozstrzygający, jak zauważane są (oraz pomijane) różnice leksykalne i gramatyczne w trakcie projektowania

wypowiedzi. Ludzie są bardziej skłonni powiedzieć „mój ruch”, gdy mają na myśli „rój much”, niż „żuma do gucia”, gdy mają na myśli „gumę do żucia”. Skłaniamy się ku prawdziwym (znanym) słowom, a nie ku słowom jedynie wymawialnym (możliwym, ale nieprawdziwym), nawet w przypadku przejęzyczeń. Niektóre błędy sugerują nam, jak muszą funkcjonować mechanizmy doboru słów: „Ta kanapa jest przygodna (przyjemna/wygodna)” czy „Piecze się w niskiej szybkości (temperaturze)”. Pomyślmy o transpozycji, która musi towarzyszyć stworzeniu następującego błędu: „przenosić zasłonę” zamiast „zanosić przesłonę”.

Dzięki genialnym eksperymentom, które wywołują takie błędy, oraz dogłębnym analizom dotyczącym tego, co się dzieje, a co nie, gdy ludzie mówią, postęp dokonuje się w modelach zorganizowanych mechanizmów odpowiedzialnych za ostateczną artykulację wiadomości, gdy już *zadecydowano*, że właśnie ta wiadomość ma zostać oznajmiona światu zewnętrznemu. Jednak kto lub co wprawia tę maszynerię w ruch? Błąd mowy jest błędem dlatego, że jest czymś innym od tego, co mówiący *chciał powiedzieć*. Jakiż przełożony *wyznacza zadanie*, względem którego oceniane są błędy takie jak wspomniane przykłady?

Co, jeśli nie centralny nadawacz sensu? Levelt kreśli obraz, „schemat mówiącego”:



Ryc. 8.1

W lewym górnym rogu funkcjonariusz, podejrzanie przypominający centralnego nadawacza sensu, pojawia się w przebraniu Konceptualizatora – uzbrojonego w ogrom wiedzy o świecie, plany i *intencje komunikacyjne* oraz zdolnego do „generowania komunikatów". Levelt ostrzega swoich czytelników, że Konceptualizator jest „reifikacją wymagającą dalszego wyjaśnienia" (s. 9), ale i tak zakłada jego istnienie, gdyż wydaje się, że nie może dalej badać procesu bez swego rodzaju niezanalizowanego szefa, dającego reszcie drużyny rozkazy do wymarszu.

Jak to działa? Kluczowy problem stanie się jaśniejszy, jeśli rozpoczniemy od karykatury. Konceptualizator decyduje się wykonać akt mowy, na przykład obrażenie swojego rozmówcy przez nieuprzejme skomentowanie rozmiaru jego stóp. Wysyła zatem rozkaz do podlegającej mu administracji, działu Public Relations (u Levelta jest to Formulator): „Powiedz temu frajerowi, że jego stopy są za duże!". Ekipa PR zabiera się do roboty. Znajdują odpowiednie słowa: zaimek dzierżawczy w drugiej osobie liczby pojedynczej, *twoje*; dobre słowo na stopy, czyli *stopy*; prawidłowa forma mnoga czasownika *być*, czyli *są*; oraz odpowiedni przysłówek i przymiotnik: *za duże*. Te słowa sprytnie łączą z odpowiednio obrażającym tonem głosu i wykonują: „Twoje stopy są za duże!".

Ale chwila, moment. Czy to aby nie za proste? Gdy Konceptualizator dał rozkaz (co Levelt nazywa *komunikatem przedwerbalnym*), to jeśli dałby go po angielsku, jak sugerowała moja karykatura, wykonałby całą ciężką pracę, zostawiając niewiele do zrobienia reszcie ekipy poza przekazaniem jej z drobnymi poprawkami. Czy więc ten przedwerbalny komunikat został przekazany w jakimś innym systemie reprezentującym lub języku? Czymkolwiek on jest, musi przekazywać podstawowe wskazówki drużynie produkcyjnej dotyczące obiektu, który ma ona zbudować i wypuścić, oraz musi być sformułowany w kategoriach dla *niej* „zrozumiałych" – nie po angielsku, ale w jakiejś wersji języka mózgowskiego czy myśleńskiego. Levelt uważa, że muszą one być sformułowane w swego rodzaju języku myśli, ale być może w języku myśli, który służy jedynie do porządkowania aktów mowy, nie do wszystkich aktywności poznawczych. Ekipa otrzymuje przedwerbalny komunikat, szczegółowy rozkaz w języku myśleńskim, że ma stworzyć wypowiedź po angielsku i wykonuje to zadanie. Daje to podwładnym trochę więcej pracy, ale jedynie przysłania czający się tu regres. W jaki sposób Konceptualizator stwierdza, których słów języka myśleńskiego użyć do wydania rozkazu? Lepiej, żeby nie było mniejszej kopii całego schematu Levelta ukrytej w module *tworzenie komunikatu*, należącym do Konceptualizatora (i tak dalej, w nieskończoność). I z pewnością nikt nie powiedział Konceptualizatorowi, co powiedzieć: to w końcu on jest centralnym nadawaczem sensu, od niego *pochodzi* znaczenie.

Jak zatem wypowiedź nabiera znaczenia? Spójrzmy na poniższy zbiór poleceń, prowadzący od wielkiej, ogólnej strategii, przez szczegółową taktykę, aż do podstawowych operacji:

(1) Obraź go!

(2) Zrób coś nieprzyjemnego, ale nie niebezpiecznego!

(3) Znieważ go!

(4) Rzuć oszczerstwo na jakiś aspekt jego wyglądu!

(5) Powiedz mu, że jego stopy są za duże!

(6) Powiedz: „Twoje stopy są za duże"!

(7) Wymów: tfoje stopɨ sɔ̃ za dɯʒe!

Z pewnością musi następować coś w rodzaju dochodzenia do końcowego aktu. Ludzka mowa jest działaniem zamierzonym: ma cele i środki, a my w jakiś sposób jako tako

przymierzamy różne warianty. Mogliśmy go popchnąć, zamiast go obrażać, lub też zdeprecjonować jego inteligencję, zamiast skupiać się na jego stopach, albo powiedzieć, cytując Fatsa Wallera: „krańce twoich kończyn dolnych są wstrętne!".

Ale czy to stopniowe przybliżanie się do ostatecznego aktu następuje w ramach biurokratycznej hierarchii dowódców dających polecenia podwładnym? W tej hierarchii służbowej pojawia się raczej sporo decyzji do podjęcia – „momenty", w których warianty są „wybierane" spośród innych, a to zachęca do uznania modelu, w którym następuje przekazywanie odpowiedzialności za szczegóły oraz w którym niższe szczeblem podmioty działające, mające swoje własne intencje, rozumieją powody, dla których dokonały wyboru. (Gdyby w ogóle nie musiały rozumieć, dlaczego robią to, co robią, tak naprawdę nie byłyby podmiotami, tylko biernymi funkcjonariuszami ze stempelkiem, pozwalającymi na przejście przez swoje biurko wszystkiemu).

Schemat Levelta wskazuje pozostałości jednego ze swoich źródeł: architekturę von Neumanna zainspirowaną refleksjami Turinga nad własnym strumieniem świadomości, która to z kolei architektura zainspirowała wiele modeli kognitywistycznych. W rozdziale 7 próbowałem przezwyciężyć opór wobec pomysłu, że ludzka świadomość jest, *trochę jak* maszyna von Neumanna, procesorem szeregowym z ciągiem określonych treści przepływających przez wąskie gardło akumulatora. Teraz muszę nieco przyhamować i zwrócić uwagę na pewne względy, pod którymi funkcjonalna architektura ludzkiej świadomości *nie* jest jak architektura maszyny von Neumanna. Jeśli porównamy schemat Levelta do typowego funkcjonowania maszyn von Neumanna emitujących słowa, możemy zobaczyć, że model Levelta zapożycza z nich ciut za dużo.

Gdy maszyna von Neumanna wyjawia, co w niej tkwi, na wyjściu podaje treść swojej jedynej, centralnej przestrzeni roboczej, akumulatora, który w każdej chwili ma całkowicie określoną treść w stałym języku arytmetyki binarnej. Elementarne „komunikaty przedwerbalne" maszyny von Neumanna wyglądają tak: 10110101 00010101 11101101. Jedną z elementarnych instrukcji w każdym języku maszynowym jest instrukcja DRUKUJ NA WYJŚCIU, która podaje obecną zawartość akumulatora (np. binarną liczbę 01100001) na ekranie lub drukarce, udostępniając zewnętrznemu użytkownikowi wyniki uzyskane w procesorze. W wersji trochę łatwiejszej dla użytkownika procedura składająca się z serii elementarnych instrukcji może najpierw przetłumaczyć liczbę binarną na system dziesiętny (np. binarne 00000110 = dziesiętne 6) lub na literę alfabetu za pomocą kodu ASCII (np. binarne 011000001 = „a", 010000001 = „A"), a następnie podać rezultat na wyjściu. Te podprogramy stanowią sedno bardziej eleganckiej obsługi wyjścia w językach programowania wyższego poziomu, takich jak Fortran, Pascal czy Lisp. Te zaś umożliwiają programiście stworzenie dalszych podprogramów do tworzenia większych komunikatów, pobierających długie szeregi liczb z pamięci i przeprowadzających je przez akumulator, tłumaczących je i podających rezultaty na ekranie bądź wydruku. Podprogram może na przykład odbyć kilka podróży do akumulatora po wartości, które wstawia w puste miejsca w ciągu tekstowym:

Szanowna/y Pani/e _____, przekroczył/a Pan/i swój stan konta o _____ zł. Miłego dnia, Pani/e _____!

– czyli w „prefabrykowanej" formule zdaniowej, która sama w sobie jest przechowywana jako szereg liczb binarnych w pamięci do momentu, aż jakaś procedura zdecyduje, że nadszedł czas na jej użycie. W ten sposób ścisła hierarchia ustalonych procedur może zamienić ciągi określonych treści w akumulatorze w wyrażenia, które człowiek może przeczytać na ekranie lub na wydruku: „Czy chcesz zapisać ten dokument?", „Skopiowano 6 plików" lub „Cześć, Wiesławie, czy chcesz zagrać w kółko i krzyżyk?".

Dwie cechy tych procesów są wspólne z modelem Levelta: (1) proces korzysta z już określonych treści jako danych wejściowych oraz (2) biurokracja – „przepływ sterowania" w żargonie informatycznym – musiała być uważnie zaprojektowana, to znaczy „podejmowanie decyzji" płynie hierarchicznie przez przydział odpowiedzialności podmiotom niższego szczebla, których opis obowiązków wskazuje, do wykonania jakich aspektów analizy środków i celów mają uprawnienie. Co ciekawe, pierwsza z tych cech – określona treść – wydaje się potwierdzać pogląd Ottona dotyczący jego własnych procesów: Istnieje pewna określona „myśl" gdzieś w centrum, czekająca na „ubranie ją w słowa". Druga wspólna cecha wydaje się jednak obca: hierarchia procedur, która wiernie odtwarza *tę właśnie myśl* w języku naturalnym, została wcześniej zaprojektowana przez *kogoś innego* – w przypadku maszyny von Neumanna przez programistę, a w przypadku czynności Formulatora Levelta prawdopodobnie przez powiązanie ewolucji i rozwoju indywidualnego. W modelu nie widać jednak twórczej roli w osądzie, którą w przekuwaniu myśli w słowa miałby odgrywać ten, kto myśli daną myśl; jest albo uzurpowana przez Konceptualizatora, który wykonuje całą kreatywną pracę przed wysłaniem rozkazu do Formulatora, albo też jest ukryta w konstrukcji Formulatora, co byłoby rezultatem jakiegoś wcześniejszego procesu konstruowania.

Jak jeszcze mogłyby być zorganizowane cele i środki? Przyjrzyjmy się przeciwnej karykaturze: pandemonium demonów słów. Mówimy w taki oto sposób. Najpierw przechodzimy do trybu wytwarzania dźwięków głosowych – włączmy róg: „Biiiiip...". Robimy to bez żadnego konkretnego powodu, ale tylko dlatego, że żaden konkretny powód, aby tego nie robić, nie przychodzi nam do głowy. Wewnętrzny „szum" pobudza w nas różne demony, które zaczynają próbować modulować róg na różne przypadkowe sposoby, zaburzając strumień. Rezultatem jest bełkot, ale jest to przynajmniej angielski bełkot (u użytkowników angielskiego):

Jaba-daba-duu-bu-bu-rum-bum-bum-pa-ram-pam-pam...

Jednak zanim któraś z tych wstydliwych rzeczy przedostanie się do świata zewnętrznego, kolejne demony, czułe na schematy w chaosie, zaczynają kształtować je w słowa, zdania, frazesy...

No i co w związku z tym?, baseball, nie wiesz, faktycznie, truskawki, zbieg okoliczności, OK? To jest ten bilet. Cóż, więc...

co pobudza demony do dalszych przypadkowych i trafnych odkryć, wzmacnianych przez okoliczności, co z kolei prowadzi do dłuższych fragmentów bardziej akceptowalnego słownictwa, aż w końcu wyłania się całe zdanie:

Wcisnę ci zęby w gardło!

Na szczęście jednak ta wypowiedź zostanie usunięta na bok, niewypowiedziana, bo w tym samym czasie (równolegle) powstawały inne konkurencyjne wypowiedzi, które mogą ujrzeć światło dzienne, w tym kilka oczywiście nieudanych, jak na przykład:

Jesteś podłym człowiekiem!

czy też:

Czytałeś ostatnio jakąś dobrą książkę?

a zwycięzca przez walkower może powiedzieć:

Twoje stopy są za duże!

Tym razem muza zawiodła naszego mówcę; żadna inteligentna odpowiedź nie dostała się do finałów, ale przynajmniej zostało wybełkotane coś mniej więcej odpowiedniego do obecnego „nastawienia" mówcy. Gdy mówca odejdzie po spotkaniu, prawdopodobnie powróci do chaotycznego turnieju, mamrocząc i rozmyślając o tym, co powinien był powiedzieć. Muza może wówczas zstąpić z czymś lepszym, a mówca będzie się tym delektował, obracając to tam i z powrotem w umyśle, wyobrażając sobie zaskoczony wyraz twarzy, który ta wypowiedź

sprowokowałaby u rozmówcy. W momencie gdy mówca wróci do domu, może dokładnie „pamiętać", jak rozłożył swojego przeciwnika na łopatki szaleńczo ciętą ripostą.

Możemy założyć, że odbywa się to w formie szybkich pokoleń „nieekonomicznego" przetwarzania równoległego, z bandą anonimowych demonów i ich pełnych nadziei konstrukcji, które nigdy nie ujrzą światła dziennego – w postaci możliwości, które są *świadomie* rozważane i odrzucane, lub w postaci ostatecznie wypowiedzianych aktów mowy, które usłyszą osoby z zewnątrz. Gdy da się im wystarczającą ilość czasu, więcej niż jeden z nich może zostać po cichu i świadomie wypróbowany, ale tak oficjalna próba jest czymś stosunkowo rzadkim, zarezerwowanym na okazje, gdy stawka jest wysoka, a powiedzenie czegoś nie tak niesie za sobą wysoką karę. W normalnym przypadku mówca nie ma możliwości pokazu przedpremierowego; on i jego publiczność dowiadują się o tym, jak wygląda wypowiedź, w tym samym czasie.

Jak przebiega sędziowanie tego turnieju między słowami? Gdy jedno słowo, zwrot lub całe zdanie pokonuje swoich przeciwników, jak odróżnia się i docenia jego odpowiedniość i stosowność do danego nastawienia? Czym *jest* nastawienie myślowe (jeśli nie jawną intencją komunikacyjną) i jak wpływa na turniej? Przecież nawet jeśli nie ma centralnego nadawacza sensu, treść musi jakoś wydostawać się z systemu – na przykład z procesów percepcyjnych – w postaci sprawozdań werbalnych.

Spójrzmy na te kwestie raz jeszcze. Problem ze skrajnością biurokratyczną polega na tym, że Konceptualizator wydaje się niepokojąco potężny, niczym homunkulus ze zbyt dużą wiedzą i odpowiedzialnością. Ten nadmiar władzy ujawnia się w postaci osobliwego problemu, jak sformułować jego wyjście, komunikat *przedwerbalny*. Jeśli już precyzuje akt mowy – jeśli jest już rodzajem mowy w myśleńskim, konkretnym poleceniu do Formulatora – praca związana z tworzeniem wypowiedzi jest skończona, zanim nasz model dojdzie do głosu. Problemem konkurencyjnego podejścia, pandemonium, jest to, że musimy wiedzieć, jak źródła treści mogą *wpłynąć* lub *ograniczyć* twórcze energie demonów słów bez *dyktowania* im czegokolwiek.

A co z procesem opisanym w rozdziale 1, wielokrotnym odpowiadaniem na pytania, który generował halucynacje w modelu gry w psychoanalizę? Zwróćmy uwagę, że wyeliminowaliśmy mądrego, freudowskiego scenopisarza snów i twórcę halucynacji, zamieniając go na proces, z którego treść *wyłaniała się* w wyniku nieprzerwanego zadawania pytań. Problemem pozostało jednak, jak pozbyć się bystrego przesłuchującego. Rozwiązanie odłożyliśmy na później. Tutaj mamy bliźniaczy problem – jak uzyskać odpowiedzi na pytania zadawane przez gorliwe stado konkurentów, na przykład: „Dlaczego nie powiemy »Twoja matka nosi wojskowe buty!«?" lub (w innym kontekście) „Dlaczego nie powiedzieć: »Wydaje mi się, że widzę poruszający się, czerwony punkt, który zmienia się w zielony podczas tego ruchu«?". *Dwa uzupełniające się problemy – może udałoby się je rozwiązać, łącząc je ze sobą?* Co, jeśli demony słów same pytają i są konkurentami, a demony treści odpowiadają i są sędziami? Pełnoprawne i wykonane intencje komunikacyjne – znaczenia – mogą wyłaniać się z quasi-ewolucyjnego procesu konstruowania aktu mowy, obejmującego współpracę, częściowo szeregową, częściowo równoległą, między różnymi podsystemami, z których żaden nie jest w stanie sam wykonać – ani zlecić wykonania – aktu mowy.

Czy taki proces naprawdę jest możliwy? Istnieją różnorodne modele tego rodzaju procesów „spełniania ograniczeń" i rzeczywiście mają one zaskakujące moce. Oprócz różnych „architektur" łączących elementy podobne do neuronów (zob. np. McClelland i Rumelhart 1986) istnieją modele bardziej abstrakcyjne. Architektura Jumbo Douglasa Hofstadtera (1983), poszukująca rozwiązań kolosów lub anagramów, ma odpowiedniego rodzaju cechy, podobnie jak idee Marvina Minsky'ego (1985) dotyczące czynników tworzących „społeczeństwo umysłu" – o czym powiemy więcej w rozdziale 9. Musimy jednak powstrzymać się od oceny, dopóki nie

zostaną stworzone i przetestowane modele bardziej szczegółowe, jasne i dotyczące bezpośrednio realizacji języka.

Wiemy wszak, że w udanym modelu realizacji języka będziemy musieli w którymś miejscu wykorzystać ewolucyjny proces tworzenia komunikatów, gdyż w innym razie pozostanie nam cud („I stał się cud"), czyli nieskończony regres nadawaczy sensów wyznaczających zadanie[74]. Wiemy również – z badań omawianych przez Levelta – że istnieją dosyć sztywne i automatyczne procesy, które w końcu przejmują kontrolę oraz determinują transformacje od gramatyki do fonetyki, a wreszcie tworzą motoryczne polecenia mowy. Te dwie karykatury stanowią ekstrema na kontinuum, od hiperbiurokratycznego po hiperchaotyczny. Rzeczywisty model Levelta – w przeciwieństwie do karykatury, której użyłem, aby jasno przedstawić kontrast – zawiera (lub może być sprawnie przekształcony tak, aby zawierał) niektóre z niebiurokratycznych cech przeciwnej karykatury: na przykład nie ma żadnej głębokiej ani strukturalnej przeszkody, która uniemożliwiałaby Formulatorowi Levelta zajęcie się mniej więcej spontanicznym (nieproszonym, nieukierunkowanym) generowaniem języka, a biorąc pod uwagę monitorującą pętlę przechodzącą przez System Rozumienia Mowy z powrotem do Konceptualizatora (zob. Ryc. 8.1), ta spontaniczna czynność *mogłaby* odgrywać swego rodzaju generującą rolę przypadającą licznym demonom słów. Między tymi dwiema karykaturami znajduje się pośrednie spektrum bardziej realistycznych konkurencyjnych modeli. Najważniejszym pytaniem jest: w jakim stopniu wzajemnie oddziałują specjaliści, którzy determinują treść oraz styl tego, co ma być powiedziane, i specjaliści, którzy „znają słowa i gramatykę"?

Na jednym skraju odpowiedź brzmi: w żadnym. Moglibyśmy pozostawić nienaruszony model Levelta, a jedynie dodać do niego model pandemonium, związany z tym, co dzieje się *wewnątrz* Konceptualizatora, aby stworzyć „komunikat przedwerbalny". W modelu Levelta oddzielenie procesów generowania komunikatów (ustalanie wymogów) oraz realizacji językowej (spełnianie wymogów) jest prawie całkowite. Kiedy pierwsza część komunikatu przedwerbalnego dociera do Formulatora, wyzwala realizację początku wypowiedzi, a gdy słowa są wybierane przez Formulatora, ogranicza to możliwe dalsze części wypowiedzi, choć istnieje minimalna *współpraca* podczas korekty wymogów. Podrzędni cieśle językowi w Formulatorze są, żeby użyć pojęcia Jerry'ego Fodora, „hermetycznie izolowani"; automatycznie robią wszystko, co się da, z otrzymywanymi rozkazami, bez żadnych „a jeśli", „oraz" czy „ale".

Na drugim skraju znajdują się modele, w których słowa i zwroty z Leksykonu, razem ze swoim brzmieniem, znaczeniem i skojarzeniami, rywalizują z konstrukcjami gramatycznymi w pandemonium, „próbując" stać się częścią wiadomości, a niektóre z nich w ten sposób mają znaczący wkład w same intencje komunikacyjne, które ostatecznie są wykonywane przez jeszcze mniejszą ich liczbę. Na tym skraju istniejące intencje komunikacyjne są tak samo efektem tych procesów, jak i ich przyczyną – pojawiają się jako wytwór, a gdy już istnieją, są dostępne jako standardy, które pozwalają ocenić *późniejsze* zrealizowanie intencji. Nie istnieje jedno źródło znaczenia, ale wiele zmieniających się źródeł, powstałych przy okazji poszukiwania odpowiednich słów. Zamiast określonej treści w określonym miejscu funkcjonalnym, czekającej na wygładzenie przez procedury niższego rzędu, nadal istnieje niecałkowicie zdeterminowane nastawienie rozproszone w mózgu i ograniczające proces tworzenia, mogące następnie zostać wykorzystane ponownie w momencie robienia poprawek i korygowania, dalej określając zadanie wyrażania treści, które uruchomiło proces tworzenia wypowiedzi. Nadal istnieje ogólny wzorzec szeregowego przechodzenia, gdzie jednocześnie koncentracja jest na jednym zagadnieniu, jednak jego granice nie są ostro zakreślone.

W modelu pandemonium sterowanie jest uzurpowane, a nie przekazywane w procesie,

który jest w dużej mierze niezaprojektowany i oportunistyczny; istnieją różnorodne źródła „decyzji" projektowych, które prowadzą do ostatecznej wypowiedzi, ale nie istnieje żaden rygorystyczny podział na odgórne rozkazy co do treści, wypływające z głębi [podmiotu], i na oddolne sugestie ich realizacji podpowiadane przez demony słów. Ten rodzaj modelu sugeruje, że aby *zachować* twórczą rolę *wyrażania* myśli (coś, co było bardzo istotne dla Ottona), musimy *porzucić* ideę, iż *podmiot* myśli zaczyna od już określonej myśli, która ma być wyrażona. Ten pomysł określonej treści również był bardzo ważny dla Ottona, lecz nie można mieć wszystkiego (w podrozdziale 3 dokładniej zanalizuję inne możliwości).

Gdzie na tym kontinuum znajduje się prawda? Jest to pytanie empiryczne, na które nie znamy jeszcze odpowiedzi[75]. Istnieją jednak zjawiska silnie wskazujące (mi), że generowanie języka wymaga pandemonium – oportunistycznych, równoległych, ewolucyjnych procesów – niemalże całkowicie. W następnym podrozdziale zajmę się pokrótce niektórymi z nich.

3. Gdy słowa chcą zostać wypowiedziane

Czegokolwiek nie chcielibyśmy powiedzieć, prawdopodobnie nie powiemy dokładnie tego.

Marvin Minsky, 1985, s. 236

Badacze sztucznej inteligencji Lawrence Birnbaum i Gregg Collins (1984) zauważyli pewną szczególną cechę freudowskich przejęzyczeń. Freud skutecznie zwrócił naszą uwagę na przejęzyczenia, które – jak twierdził – nie są przypadkowe i bez znaczenia, ale są głęboko sensowne: nieświadomie *zamierzone* wstawki do materiału dyskursywnego, wstawki, które pośrednio lub częściowo spełniają stłumione cele komunikacyjne nadawcy. To standardowe twierdzenie freudowskie często jest zaciekle odrzucane przez sceptyków, ale jest coś zagadkowego w jego zastosowaniu w konkretnych przypadkach, które nie ma nic wspólnego z opinią o ciemnych motywach seksualnych Freuda, z kompleksem Edypa czy życzeniem śmierci. Freud omówił przypadek, w którym pewien mężczyzna powiedział:

„Szanowni Państwo, proponuję, abyśmy *czknęli* za zdrowie naszego naczelnika".

(Po niemiecku – czyli w języku tego przykładu – słowo oznaczające „mieć czkawkę", *aufzustossen*, zastąpiło słowo oznaczające „pić", *anzustossen*).

W swoim wyjaśnieniu Freud zakłada, że to przejęzyczenie jest manifestacją nieuświadomionego celu ze strony nadawcy, chcącego wyśmiać lub obrazić swojego przełożonego, który to cel jest stłumiony przez społeczny i polityczny obowiązek, aby go szanować. Jednak nie można racjonalnie oczekiwać, że intencja nadawcy, chcącego ośmieszyć swojego przełożonego, wywołała plan zakładający użycie słowa „czkać": *a priori* istnieją setki słów i zdań, które mogą bardziej przekonująco posłużyć do obrażenia lub ośmieszenia kogoś. Nie jest możliwe, aby dało się słusznie przewidzieć, że cel ośmieszenia bądź obrażenia przełożonego zostanie spełniony przez wymówienie słowa „czkać" z dokładnie tego samego powodu, dla którego nieprzekonujący jest w ogóle wybór tego słowa w celu obrazy.

Birnbaum i Collins uważają, że jedyny proces, który mógłby wyjaśnić częstotliwość szczęśliwych freudowskich przejęzyczeń, to „planowanie oportunistyczne".

Przykłady takie jak powyższy zdają się zatem wskazywać, że cele *same w sobie* są czynnymi podmiotami poznawczymi, zdolnymi do rozkazywania potrzebnym zasobom poznawczym, potrzebnym do rozpoznania okazji ich zaspokojenia, oraz zasobom behawioralnym potrzebnym do wykorzystania tych okazji. [Birnbaum i Collins 1984, s. 125]

Przejęzyczenia freudowskie przykuwają uwagę, gdyż jednocześnie wydają się i nie

wydają pomyłkami, ale fakt (jeśli nim jest), że zaspokajają *nieuświadomione* cele, nie sprawia, iż są trudniejsze do wytłumaczenia niż inny dobór słów, który spełnia różne funkcje (lub cele) jednocześnie. Jest niemal równie trudno wyobrazić sobie, jak dowcipy oraz inne formy celowego humoru werbalnego mogą być wynikiem nieoportunistycznego, izolowanego planowania i realizacji. Jeśli ktoś ma plan zaprojektowania żartu – szczegółowy plan, który rzeczywiście działa – wielu komików zapłaciłoby za niego wiele pieniędzy[76].

Jeśli Birnbaum i Collins mają rację, twórcze użycie języka może powstać tylko na drodze równoległego procesu, w którym wiele *celów* jest jednocześnie w stanie pogotowia i przeszukuje środki wyrazu. A co jeśli same środki wyrazu były w tym samym czasie w pogotowiu i czekały na okazje przyłączenia? Uczymy się naszego słownictwa od naszej kultury; wyrazy i zdania są najbardziej istotnymi cechami fenotypowymi – widzialnymi ciałami – memów, które nas okupują, i chyba nie mógłby istnieć lepszy ośrodek replikacji memów niż system realizacji języka, gdzie nadzorujący biurokraci częściowo abdykowali, przekazując sporą dozę kontroli samym słowom, które faktycznie walczą między sobą o publiczną ekspresję.

Wiemy, że czasem mówimy coś głównie dlatego, że podoba nam się, jak to brzmi, a nie to, co znaczy. Nowe potoczne wyrażenia rozchodzą się w społecznościach, dostając się do wypowiedzi niemal każdej osoby, nawet tej, która stara się im sprzeciwić. Niewiele osób używających nowego słowa celowo i świadomie wypełnia maksymę nauczycieli: „Użyj słowa trzy razy, a stanie się twoje!". A jeśli chodzi o wyższy poziom łączenia słów, to całe zdania przypadają nam do gustu z powodu ich brzmienia, raczej niezależnie od tego, czy zgadzają się z ustalonymi już wcześniej wymogami co do sądów. Jedną z najczęściej cytowanych wypowiedzi Abrahama Lincolna jest:

Można oszukiwać wszystkich ludzi przez jakiś czas lub niektórych ludzi przez cały czas, ale nie można ciągle oszukiwać wszystkich[77].

Co miał na myśli Lincoln? Nauczyciele logiki lubią zwracać uwagę, że w tym zdaniu występuje „wieloznaczność zasięgowa". Czy Lincoln chciał stwierdzić, że istnieją matołki, które zawsze można oszukać, czy że przy każdej okazji ten czy inny osobnik zostanie oszukany – ale nie będą to zawsze ci sami ludzie? Logicznie są to zupełnie różne sądy.

Porównaj:

„Ktoś zawsze wygrywa na loterii".

„To musiało być ustawione!"

„Nie o to mi chodziło".

Która z interpretacji była zamierzona przez Lincolna? Może żadna! Jaka jest szansa, że Lincoln nigdy nie zauważył wieloznaczności zasięgowej i nigdy tak naprawdę nie miał jednej intencji komunikacyjnej, a nie „drugiej"? Być może po prostu zdanie brzmiało mu tak dobrze, gdy po raz pierwszy je sformułował, że nigdy nie dostrzegł dwuznaczności i *nigdy nie miał wcześniejszej intencji komunikacyjnej* – poza intencją, aby powiedzieć coś konkretnego i w dobrym rytmie na ogólny temat oszukiwania ludzi. Ludzie naprawdę tak mówią, nawet tak wielcy nadawacze sensu jak Lincoln.

Prozatorka Patricia Hampl w swoim wnikliwym eseju *The Lax Habits of the Free Imagination* (Rozluźnione nawyki wolnej wyobraźni) pisze, jak tworzy opowiadania:

Każde opowiadanie ma historię. Ta tajemnicza historia, która ma niewielką szansę na opowiedzenie, jest historią jego stworzenia. Być może „historia opowiadania" *nie może* nigdy zostać opowiedziana, gdyż ukończona praca wchłania swoją własną historię, czyni ją przestarzałą łupiną. [Hampl 1989, s. 37]

Ukończone dzieło, powiada pisarka, jest wystawione na interpretacje krytyków niczym podstępnie wymyślny artefakt, kryjący mnóstwo wyszukanych intencji autorskich. Gdy jednak

napotyka taką hipotezę na temat własnej twórczości, jest jej wstyd:

„Hampl" miała niewiele cennych intencji, oprócz tego, by niczym szarlatan, którym nagle się poczułam, zwędzić to, co leżało luzem na stole i co pasowało do moich chwilowych celów. Co gorsza, „cele" były mgliste, niespójne, nieostateczne, pod presją. Kto – lub co – tworzyło tę presję? Nie umiem powiedzieć. [Hampl 1989, s. 37]

Jak więc to robi? Proponuje maksymę: „Po prostu nie przerywaj mówienia – mamrotanie jest w porządku". W końcu mamrotanie przyjmuje formy, które spotykają się z aprobatą autorki. Czy jest możliwe, że procesy, które wykrywa Hampl, działające na dużą skalę w przypadku jej twórczego pisania, są po prostu wzmocnieniem bardziej ukrytego i szybkiego procesu, który na co dzień prowadzi do twórczego mówienia?

To kuszące podobieństwo nie zakłada jedynie procesu, ale również następujące po nim postawę czy reakcję. Zapał Hampl do wyznań kontrastuje z bardziej normalną – i niekoniecznie nieszczerą – reakcją autorów na przyjazne interpretacje czytelników: owi autorzy zdają się na przypisywane im intencje, a nawet chętnie się na ich temat wypowiadają, w duchu: „No, chyba właśnie o to mi od samego początku chodziło!". A czemu nie? Czy jest coś samosprzecznego w myśli, że pewien ruch, który ktoś właśnie wykonał (w szachach, w życiu, w pisaniu), jest tak naprawdę mądrzejszy, niż początkowo sądziliśmy? (Dalsze przemyślenia na ten temat znajdziesz w Eco 1990).

Jak to ujął E.M. Foster: „Skąd mam wiedzieć, co myślę, zanim zobaczę, co powiedziałem?". Często rzeczywiście odkrywamy, co uważamy (a więc – co mamy na myśli), zastanawiając się nad tym, co widzimy, że mówimy – i nie poprawiając tego. Jedziemy więc, przynajmniej w takich sytuacjach, na tym samym wózku co nasi zewnętrzni krytycy i interpretatorzy, znajdując fragment tekstu i analizując go najlepiej, jak potrafimy. To, co powiedzieliśmy, nadaje temu swego rodzaju osobiste przekonanie, a przynajmniej przypuszczenie, autentyczności. *Prawdopodobnie*, jeśli to powiedziałem (a słyszałem siebie, gdy to mówiłem, i nie słyszałem siebie spieszącego z poprawkami), to miałem na myśli i prawdopodobnie znaczy to, co wydaje się znaczyć – dla mnie.

Życie Bertranda Russella daje przykład:

Gdy dwóch gości wyszło, było już późno i Russell został sam z Lady Ottoline. Siedzieli przy kominku i rozmawiali do czwartej nad ranem. Russell, odnotowując to wydarzenie kilka dni później, napisał: „Nie wiedziałem, że cię kocham, dopóki nie usłyszałem siebie, gdy ci to mówiłem – w pierwszej chwili pomyślałem »O Boże, co ja powiedziałem?«, ale potem wiedziałem, że to prawda". [Clark 1975, s. 176]

Co jednak z pozostałymi okazjami, w których nie mamy tego rodzaju poczucia *odkrycia* autointerpretacji? Możemy przypuszczać, że w tych normalnych przypadkach mamy swego rodzaju bezpośredni, uprzywilejowany i wcześniejszy dostęp do tego, co mamy na myśli, tylko dlatego, że to my sami jesteśmy nadawaczami sensu, *źródłem i przyczyną* znaczenia słów, które *my* wypowiadamy, jednak takie założenie wymaga dowodu potwierdzającego, a nie tylko odwołania do tradycji. Z tego względu, że równie dobrze możemy nie mieć poczucia odkrycia w tych sytuacjach dlatego, iż to, co mamy na myśli, jest dla nas tak oczywiste. Nie potrzeba „uprzywilejowanego dostępu", aby wyczuć, że gdy mówię „Podaj, proszę, sól" przy stole, proszę o sól.

Kiedyś uważałem, że nie da się zastąpić centralnego nadawacza sensu, i myślałem, że znalazłem dla niego bezpieczną przystań. W książce *Content and Consciousness* przekonywałem, że musi istnieć funkcjonalnie wyraźna linia (którą nazwałem linią świadomości [*awarness line*]) oddzielająca przedświadome ustalenie intencji komunikacyjnych od ich późniejszego wykonania. Przebieg tej linii w mózgu może być potwornie zagmatwany, w sensie anatomicznym, ale musi

ona istnieć, ze względów logicznych, niczym podział dysfunkcji na dwie kategorie. Błędy mogą występować gdziekolwiek w całym systemie, jednak każdy z nich musi się znaleźć – z geometrycznej konieczności – po jednej stronie linii lub po drugiej. Jeśli znalazł się po wykonawczej stronie linii, staje się błędem *ekspresji* (możliwym do naprawienia), takim jak przejęzyczenie, malapropizm czy inny błąd językowy. Jeśli znalazł się po wewnętrznej czy wyższej stronie linii, *zmieniał to, co miało zostać wyrażone* („komunikat przedwerbalny" w modelu Levelta). Znaczenie jest ustalone w tym miejscu; stąd przychodzi znaczenie. Myślałem, że musi istnieć miejsce, z którego pochodzi znaczenie, ponieważ *coś* musi wyznaczać standard, według którego „informacja zwrotna" może zarejestrować błąd w wykonaniu.

Moim błędem było popadnięcie w dokładnie tę samą wieloznaczność zasięgową, która nęka interpretację powiedzenia Abrahama Lincolna. Rzeczywiście, przy każdej okazji musi istnieć coś, co jest dla niej standardem, zgodnie z którym poprawiony zostaje każdy „błąd", jednak nie musi to być za każdym razem ta sama, pojedyncza rzecz – nawet *w trakcie trwania* jednego aktu mowy. Nie musi istnieć ustalona (ani zagmatwana) linia, która wyznacza to rozróżnienie. Tak naprawdę, jak widzieliśmy w rozdziale 5, rozróżnienie modyfikacji *przed przeżyciem zmieniających świadomość* i modyfikacji po przeżyciu, które prowadzą do *błędnego relacjonowania czy raportowania świadomości*, jest nieokreślone. Niekiedy badani starają się ponowne przyjrzeć czy zmienić to, co uznawali za prawdę, a czasami nie. Czasem, gdy wprowadzają korekty, zredagowana narracja nie jest bliższa „prawdy" czy tego, „co *naprawdę* mieli na myśli", niż wersja pierwotna. Jak zauważyliśmy wcześniej, to, gdzie kończy się redagowanie przed wydaniem, a zaczyna tworzenie erraty po wydaniu, jest kwestią, którą można rozstrzygnąć tylko w sposób arbitralny. Gdy zadajemy badanemu pytanie, czy pewne publiczne wyznanie w sposób adekwatny oddaje wewnętrzną prawdę o tym, co właśnie przeżył, osoba badana wcale nie ma lepszego prawa do oceny niż my, zewnętrzni obserwatorzy. (Zobacz również Dennett 1990d).

Oto kolejny punkt widzenia na to zjawisko. Kiedy trwa tworzenie wyrażenia werbalnego, na początku istnieje odległość, która musi zostać zniwelowana: moglibyśmy ją nazwać „odległością niedopasowania w przestrzeni semantycznej", między treścią, która ma zostać wypowiedziana, a różnymi potencjalnymi wyrażeniami werbalnymi, które początkowo się pojawiają. (Wedle moich wcześniejszych poglądów traktowałem to jako problem prostej „korekty na podstawie informacji zwrotnej", ze standardem w formie *ustalonego punktu*, względem którego potencjalne wypowiedzi ocenia się, eliminuje bądź ulepsza). Proces „uzgadniania" (*back-and-forth*), który zmniejsza odległość, jest rodzajem procesu informacji zwrotnej, ale dla treści do wyrażenia jest równie dobrze możliwe, że zostanie ona skorygowana pod kątem pewnego możliwego wyrażenia, jak i że potencjalne wyrażenie zostanie zmienione, aby lepiej pasowało do wyrażanej treści. Tak oto najlepiej dostępne czy osiągalne słowa lub zdania mogłyby tak naprawdę *zmienić treść* przeżycia (jeśli rozumiemy przeżycie jako coś, co ostatecznie jest relacjonowane – ustalone zdarzenie w heterofenomenologicznym świecie osoby badanej)[78].

Jeśli nasza jedność jako nadawaczy sensu nie jest lepiej gwarantowana, to w zasadzie powinno być możliwe zaburzenie jej czasami. Oto dwie sytuacje, w których bodaj coś takiego obserwujemy.

Zostałem kiedyś namówiony na odegranie roli sędziego na pierwszej bazie w meczu baseballowym – obowiązek ten był dla mnie nowością. W kluczowym momencie meczu (druga połowa dziewiątej rundy, dwóch zawodników wyautowanych, zawodnik na trzeciej bazie biegnący po remis) moim zadaniem było zadecydować o statusie pałkarza biegnącego do pierwszej bazy. Byłem blisko wpadki, ale okazało się, że stanowczo szarpię kciuk w górę –

sygnalizując OUT – krzycząc jednocześnie: „W polu!”. W wyniku powstałego zamieszania zostałem poproszony o powiedzenie, co tak naprawdę miałem na myśli. Nie byłem w stanie tego zrobić, przynajmniej z żadnej uprzywilejowanej pozycji. W końcu zdecydowałem (w duchu), że skoro nie mam doświadczenia w sygnalizowaniu rękoma, ale za to jestem kompetentnym mówcą, to mój akt wokalny powinien zostać uznany za prawdziwy, jednak takiej oceny mógł dokonać ktokolwiek inny. (Bardzo chciałbym poznać anegdoty o innych sytuacjach, w których ludzie nie wiedzieli, który z dwóch aktów mowy chcieli wykonać).

W kontekście eksperymentalnym psycholog Tony Marcel (1993) odkrył jeszcze bardziej spektakularny przypadek. Badany, cierpiący na ślepowidzenie (o którym powiem więcej w rozdziale 11), został poproszony o *powiedzenie*, kiedy myślał, że widzi błysk światła, jednak dano mu dziwne instrukcje, jak ma to zrobić. Poinstruowano go, aby wyraził ten jeden akt mowy na trzy różne sposoby naraz (nie po kolei, ale też niekoniecznie „unisono”):

(1) powiedzenie „Tak”,

(2) wciśnięcie przycisku (przycisku TAK),

(3) mrugnięcie TAK.

Zdumiewające jest to, że badany nie zawsze wykonywał wszystkie trzy polecenia razem. Czasami mrugał, ale nie mówił „tak” lub nie wciskał przycisku, i tak dalej. Trzech różnych odpowiedzi nie można było łatwo zhierarchizować ani pod kątem wierności intencji, ani trafności. Innymi słowy, gdy pojawiały się takie niezgodności między czynnościami, badany nie miał żadnego wzorca, według którego mógłby stwierdzić, który akt akceptuje, a który można uznać za przejęzyczenie, błąd palca czy powieki. Przyszłość pokaże, czy podobne wyniki mogą być otrzymane w innych warunkach, u innych badanych, zdrowych lub nie, lecz inne patologie również wskazują na model realizacji mowy, w którym werbalizacja *może* zostać uruchomiona bez polecenia centralnego nadawacza sensu. Jeśli cierpisz na jedną z tych patologii, „umysł pojechał ci na wakacje, ale usta mają nadgodziny”, jak śpiewał Mose Allison.

Afazja jest utratą lub upośledzeniem umiejętności mówienia, a kilka jej odmian jest dość powszechnych i szeroko zbadanych przez neurologów i językoznawców. W najbardziej powszechnej odmianie, afazji Broki, pacjent jest dotkliwie świadomy problemu i z narastającą frustracją próbuje znaleźć słowa, które ma na końcu języka. W afazji Broki istnienie zablokowanych intencji komunikacyjnych jest dla pacjenta boleśnie jasne. Jednak w stosunkowo rzadkiej odmianie afazji, afazji żargonowej, wydaje się, że pacjenci nie przeżywają w ogóle żadnego stresu związanego ze swoim werbalnym deficytem[79]. Mimo że są inteligentni i zupełnie niepsychotyczni czy z demencją, wydają się całkowicie zadowoleni z wypowiedzi takich jak te (wzięte z dwóch przypadków opisanych przez Kinsbourne’a i Warringtona [1963]):

Przypadek 1:

Jak się dzisiaj miewasz?

– Plotkowanie OK i lordowie i krykiet i Anglia i Szkocja walczy. Nie wiem. Nadciśnienie i dwa wygrane krykiet, kręgle, mruganie i łapanie, biedaki, anulowanie, ręka i kłótnia, kończąc kręgle.

Co znaczy „najważniejsze jest bezpieczeństwo”?

– Patrzeć i wiedzieć i Richmond Road w szczególności i patrzeć ruch uliczny i wahanie prawo i spacerowanie, bardzo dobry powód, być może, zebry może te, samochód osobowy i światła drogowe.

Przypadek 2:

Czy pracowałeś w biurze?

– Tak, pracowałem w biurze.

A jaka to była firma?

– O, jako kierownik tego, a skarga była taka, żeby przedyskutować odcienie dotyczące tego, jakiego były rodzaju, jak były pisane, i trzymane od różnych... triku... trikula, żeby zabrać mnie z przypisanych konwenii... przepraszam...

– Chce dać jedno subiektywne powołanie, aby utrzymać powołanie idealnego zapłodnienia rodzemstwa.

– Jej zwykłym marszczowaniem była kropka.

poproszony o zidentyfikowanie pilnika do paznokci:

– To jest nóż, ogon noża, nieświeży, nieświeży nóż.

i nożyczek:

– Gaje – to jest gaje – to nie jest tak naprawdę gaje – dwa gaje zawierające grzebień – nie, nie grzebień – dwa gaje, jeśli komendant nie jest teraz.

Dziwnie podobna przypadłość, a o wiele bardziej powszechna, to *konfabulacja*. W rozdziale 4 sugerowałem, że normalni ludzie mogą często konfabulować o szczegółach własnych przeżyć, gdyż są podatni na zgadywanie, nie uświadamiając sobie tego, i mylą teoretyzowanie z obserwacjami. Patologiczna konfabulacja jest nieświadomą fikcją zupełnie innego kalibru. Często w przypadku uszkodzenia mózgu, zwłaszcza gdy pacjenci doświadczają potwornej utraty pamięci – jak w zespole Korsakowa (typowe następstwo ostrego alkoholizmu) – mimo wszystko nadal gawędzą o swoim życiu i minionych historiach, opowiadając rzeczy całkowicie nieprawdziwe, a w przypadku ostrej amnezji zmyślają nawet wydarzenia sprzed paru minut.

Słowa będące rezultatem konfabulacji brzmią zupełnie normalnie. Często wydaje się, że są jak mało konkretna, schematyczna pogaduszka, którą w barze uważa się za rozmowę: „O, tak, ja i moja żona – mieszkamy w tym samym domu od trzydziestu lat – jeździliśmy kiedyś na Coney Island, no i wiesz, siadaliśmy na plaży – *uwielbiałem* siadać na plaży i patrzeć na młodych ludzi, a to było przed wypadkiem...” – z wyjątkiem tego, że wszystko jest zmyślone. Żona tego człowieka mogła umrzeć wiele lat temu i nigdy nie być bliżej niż sto kilometrów od Coney Island, a przy tym mogli często zmieniać mieszkania. Niewtajemniczony słuchacz może być często zupełnie nieświadomy, że rozmawia z osobą konfabulującą, gdyż jej opowieści i gotowe odpowiedzi na pytanie są tak naturalne i „szczere”.

Osoby konfabulujące nie mają pojęcia, że to wszystko zmyślają, a cierpiący na afazję żargonową są nieświadomi faktu, że wylewają z siebie nic nieznaczące strumienie słów. Te szokujące anomalie są przykładami *anosognozji*, czyli nieumiejętności przyznania czy rozpoznania deficytu. Istnieją inne rodzaje takiego braku autokontroli i w rozdziale 11 zobaczymy, co mogą nam one powiedzieć o funkcjonalnej architekturze świadomości. W międzyczasie możemy zauważyć, że maszyneria mózgowa jest zdolna do skonstruowania pozornych aktów mowy przy braku jakiejkolwiek spójnej wskazówki z góry[80].

Patologia, czy to chwilowe nadwyrężenie wywołane sprytnymi eksperymentami, czy bardziej stały defekt spowodowany chorobą lub mechanicznym uszkodzeniem mózgu, dostarcza mnóstwo wskazówek na temat organizacji tej maszynerii. Te zjawiska oznaczają dla mnie, że nasza druga karykatura, pandemonium, jest bliższa prawdy niż dostojniejszy, biurokratyczny model, należy jednak tę kwestię poddać odpowiednim testom empirycznym. Nie twierdzę, że nie jest możliwe, aby jakiś w dużej mierze zbiurokratyzowany model mógł wyjaśnić te patologie, ale że nie wydają się one naturalnymi usterkami takiego systemu. W załączniku B, dla naukowców, wspomnę o kierunkach badań, które mogłyby pomóc potwierdzić to przeczucie lub je obalić.

W tym rozdziale naszkicowałem – ale z pewnością nie udowodniłem – sposób, w jaki burza werbalnych wytworów wynurzająca się z tysięcy demonów słownych w chwilowych koalicjach może przejawiać jedność, jedność ewoluującej, najlepszej interpretacji, która sprawia,

że wytwory te wydają się, *jak gdyby* były wykonanymi intencjami Konceptualizatora – i rzeczywiście nimi są, ale nie intencjami Konceptualizatora *wewnętrznego*, który jest właściwą częścią systemu realizującego język, lecz intencjami Konceptualizatora globalnego, czyli osoby, której system produkujący język jest sam w sobie właściwą częścią.

Ten pomysł z początku może się wydawać obcy, jednak nie powinien nas zaskakiwać. W biologii nauczyliśmy się opierać pokusie wyjaśniania *projektu w organizmach* twierdzeniem, że istnieje wspaniała inteligencja, która zań odpowiada. W psychologii nauczyliśmy się opierać pokusie wyjaśniania *widzenia*, mówiąc, że to tak, jak byśmy byli wewnętrznymi obserwatorami ekranu, gdyż ten wewnętrzny obserwator zajmuje się wszystkim – jedyną rzeczą między takim homunkulusem a oczami jest rodzaj kabla do telewizora. Musimy wytworzyć taki sam opór przeciwko pokusie, aby wyjaśniać *czynność* jako powstającą z rozkazów wewnętrznego zleceniodawcy czynności, który zna za dużo szczegółowych wymagań wobec tych czynności. Jak zwykle sposobem na pozbycie się zbyt dużej inteligencji w naszej teorii jest zamienienie jej na ostatecznie mechaniczny materiał półniezależnych półinteligencji działających wspólnie.

Ta kwestia nie ogranicza się tylko do czynności realizowania mowy; możemy ją stosować w ogóle do czynności intencjonalnych. (Pears [1984] rozwija podobną koncepcję). Wbrew pozorom fenomenologia pomaga nam zobaczyć, że tak właśnie jest. Choć czasem mamy świadomość wykonywania rozbudowanego rozumowania praktycznego prowadzącego do wniosku, co, zważywszy na wszystko, powinniśmy zrobić, po czym następuje świadoma decyzja, aby zrobić właśnie to, a potem właśnie to robimy, to są to stosunkowo rzadkie przeżycia. Większość naszych czynności intencjonalnych jest wykonywana bez takiego wstępu, co jest dobre, bo nie ma na to czasu. Typowa pułapka to założenie, że te stosunkowo rzadkie przypadki świadomego, praktycznego rozumowania są dobrym modelem dla reszty, czyli dla przypadków, w których nasze intencjonalne czynności wyłaniają się z procesów, do których nie mamy żadnego dostępu. Nasze czynności zwykle nas zadowalają; stwierdzamy, że na ogół są spójne i że odpowiednio szybko pomagają w naszych przedsięwzięciach o tyle, o ile je rozumiemy. Bezpiecznie zatem zakładamy, że są wynikiem procesów niezawodnie czułych na cele i środki. Innymi słowy, są racjonalne, w *jednym* znaczeniu tego słowa (Dennett 1987a, 1991a). Nie oznacza to jednak, że są racjonalne w węższym znaczeniu: jako wyniki szeregowego rozumowania. Nie musimy wyjaśniać procesów na obraz wewnętrznego mędrka, wyciągającego wnioski, podejmującego decyzje, który metodycznie dopasowuje środki do celów, a następnie nakazuje konkretną czynność; widzieliśmy w zarysie, jak inny rodzaj procesu może kierować mową oraz naszymi innymi czynnościami intencjonalnymi.

Powoli porzucamy nasze złe nawyki myślowe i zastępujemy je innymi nawykami. Upadek centralnego nadawacza sensu jest w ogóle upadkiem centralnego twórcy intencji, jednak szef nadal ma się dobrze w innych przebraniach. W rozdziale 10 spotkamy go w rolach obserwatora i reportera i będziemy musieli zacząć inaczej pojmować to, co się dzieje, ale najpierw musimy się upewnić, że bezpieczne są podwaliny naszych nowych nawyków myślowych, bliżej badając ich naukowe szczegóły.

Rozdział 9

Architektura ludzkiego umysłu

1. Gdzie jesteśmy?

Najtrudniejsze za nami, lecz nadal pozostało nam dużo pracy. Zakończyliśmy właśnie najbardziej wymagające ćwiczenia rozciągające wyobraźni i jesteśmy gotowi poddać próbom naszą nową perspektywę. Po drodze musieliśmy pozostawić kilka nierozwiązanych tematów i przełknąć sporo jedynie markowanych argumentów. Teraz należy dotrzymać obietnic, złożyć podziękowania i przeprowadzić porównania. Rozwijana przeze mnie teoria zawiera elementy zapożyczone od wielu myślicieli. Czasem celowo pomijałem to, co owi myśliciele uważają za najlepsze elementy swoich teorii, oraz mieszałem z tym idee zapożyczone z „wrogich" obozów, ale zatrzymałem dla siebie te liczne szczegóły, aby nie zakłócić jasności i wyrazistości. Być może sprawiło to, że niektórzy poważni specjaliści od modelowania mózgu wiją się z frustracji, lecz nie sposób było inaczej doprowadzić różnego rodzaju czytelników do tej samej mety. Natomiast w tej chwili możemy się już cofnąć i przejrzeć najważniejsze szczegóły. Przecież w końcu po to pracowicie kreślę nową perspektywę, aby można było w innym świetle zobaczyć zjawiska i spory. Rozejrzyjmy się zatem.

Oto moja dotychczasowa teoria w miniaturce:

Nie istnieje jeden definitywny „strumień świadomości", gdyż nie ma głównej siedziby czy teatru kartezjańskiego, gdzie „wszystko łączy się w jedno" na użytek centralnego nadawacza sensu. Zamiast takiego pojedynczego strumienia (choćby i szerokiego) istnieje wiele kanałów, w których specjalistyczne obwody próbują, w równoległych pandemoniach, zajmować się swoimi obowiązkami, tworząc po drodze wielokrotne szkice. Większość tych fragmentarycznych szkiców „narracji" odgrywa krótkotrwałe role, modulując bieżące czynności, jednak niektóre dostają awans i otrzymują kolejne role funkcjonalne, raz za razem, dzięki działaniu maszyny wirtualnej w mózgu. Szeregowość owej maszyny (jej „von neumannowski" charakter) nie jest cechą zainstalowaną na stałe, a raczej rezultatem następujących po sobie zgrupowań tych specjalistów.

Podstawowi specjaliści należą do naszego dziedzictwa zwierzęcego. Nie powstali po to, aby wykonywać poszczególne czynności ludzkie, jak czytanie czy pisanie, ale takie jak uchylanie się, unikanie drapieżników, rozpoznawanie twarzy, chwytanie, rzucanie, zbieranie jagód i inne kluczowe umiejętności. Często są oportunistycznie werbowani do nowych ról, do których mniej więcej pasują ich przyrodzone talenty. Rezultatem nie jest chaos tylko dlatego, że tendencje narzucone tym czynnościom same w sobie są wynikiem konstrukcji. Część tej konstrukcji jest wrodzona i dzielimy ją z innymi zwierzętami. Konstrukcja ta jest jednak udoskonalona, a czasem nawet zupełnie zmarginalizowana przez mikronawyki myślowe rozwijające się w indywidualnych, częściowo specyficznych rezultatach autoeksploracji, a częściowo w gotowych wytworach kultury. Tysiące memów, zrodzonych głównie przez język, ale również przez bezsłowne „obrazy" i inne struktury danych, zamieszkują indywidualny mózg, kształtując jego tendencje i w ten sposób przemieniając go w umysł.

Ta teoria jest wystarczająco nowa, aby początkowo trudno było ją pojąć, lecz opiera się na modelach rozwiniętych w psychologii, neurobiologii, sztucznej inteligencji, antropologii –

oraz filozofii. Na taki niepohamowany eklektyzm badaczy z dziedzin, od których zapożycza, z reguły krzywo patrzą. Jako częsty intruz na tych polach, przyzwyczaiłem się do braku szacunku, wyrażanego przez niektórych badaczy wobec swoich współpracowników z innych dyscyplin. „Daniel, dlaczego – pytają informatycy parający się sztuczną inteligencją – marnujesz czas, konsultując się z tymi neuronaukowcami? Nie interesuje ich »przetwarzanie informacji«, martwią się tylko tym, *gdzie* ono zachodzi i które neuroprzekaźniki są wymagane oraz innymi nudnymi faktami, ale nie mają pojęcia o obliczeniowych wymaganiach wyższych funkcji kognitywnych". „Dlaczego – pytają neuronaukowcy – marnujesz czas na fantazje o sztucznej inteligencji? Wymyślają sobie urządzenia, na jakie mają ochotę, i mówią niewybaczalne bzdury o mózgu". W międzyczasie psychologowie poznawczy są oskarżani o wymyślanie modeli pozbawionych zarówno wiarygodności biologicznej, jak i udowodnionych mocy obliczeniowych; antropologowie nie potrafią nawet rozpoznać modelu, a filozofowie, jak wszyscy świetnie wiemy, piorą cudze brudy, ostrzegając przed zamieszaniem, które sami wywołali, w sferze pozbawionej zarówno danych, jak i empirycznie sprawdzalnych teorii. Nic dziwnego, że świadomość nadal jest zagadką, jeśli zajmuje się nią tylu idiotów.

Wszystkie te zarzuty są prawdziwe, ale idiotów na razie nie poznałem. Naukowcy, z których pracy zapożyczyłem poszczególne elementy, są, moim zdaniem, bardzo mądrzy – a nawet genialni, a przy tym aroganccy i niecierpliwi, bo te cechy często towarzyszą geniuszowi – lecz nie mają wielkich perspektyw i planów, a jedynie starają się przyczynić do rozwiązania trudnych problemów, podążając każdym skrótem, który *oni sami* dostrzegą, jednocześnie potępiając skróty innych ludzi. Nikt nie może jasno postrzegać wszystkich problemów i detali, nawet ja, i wszyscy muszą mamrotać, zgadywać i zapoznawać ważne elementy problemu.

Na przykład jednym z zawodowych niebezpieczeństw neuronauki wydaje się tendencja do postrzegania świadomości jako *końca procesu*. (Jest to jak zapominanie, że ostatecznym wytworem jabłoni nie są jabłka – to więcej jabłoni). Oczywiście dopiero od niedawna neuronaukowcy w ogóle pozwalają sobie myśleć o świadomości i tylko niewielu odważnych teoretyków odważa się mówić oficjalnie na temat swoich przekonań. Jak zażartował ostatnio badacz wzroku Béla Julesz, może się to upiec tylko wtedy, gdy masz siwe włosy – i Nagrodę Nobla! Oto na przykład hipoteza, jaką zaryzykowali Francis Crick i Christof Koch:

Sugerujemy, że jedną z funkcji świadomości jest przedstawianie wyników różnych ukrytych obliczeń oraz że zakłada to mechanizm uwagi, który czasowo wiąże odpowiednie neurony przez synchronizację ich iglic w oscylacjach o częstotliwości 40 Hz. [Crick i Koch 1990, s. 272]

Więc funkcją świadomości jest *przedstawianie wyników różnych ukrytych obliczeń* – ale komu? Królowej? Crick i Koch nie zadają następnie trudnego pytania: *A co potem?* („I staje się cud"?) Gdy tylko ich teoria doprowadziła ich do czegoś, co uważają za zaklęty krąg świadomości, urywa się. Nie rozwiązuje na przykład problemów, którymi zajęliśmy się w rozdziałach od 5 do 8, dotyczących zdradliwej ścieżki wiodącej od (przypuszczalnej) świadomości do zachowania, w tym głównie raportami introspekcyjnymi.

Natomiast modele umysłu proponowane przez psychologię poznawczą i badaczy sztucznej inteligencji prawie nigdy nie ciepią na *ten* defekt (zob. np. Shallice 1972, 1978; Johnson-Laird 1983, 1988; Newell 1990). Zwykle zakładają „przestrzeń roboczą" czy „pamięć roboczą", która zastępuje teatr kartezjański, a modele pokazują, jak wyniki przeprowadzonych obliczeń wpływają na kolejne obliczenia, które sterują zachowaniem, przekazują informacje do relacji werbalnych, zawracają rekurencyjnie, aby dostarczyć nowe informacje wejściowe do pamięci roboczej, i tak dalej. Jednak te modele nie mówią, gdzie lub jak pamięć robocza miałaby być zlokalizowana w mózgu, i są tak zajęte *pracą* w przestrzeni roboczej, że nie mają czasu na

„zabawę" – nie ma tam śladu żadnego rodzaju *rozkoszowania się* fenomenologią, która wydaje się tak ważną cechą ludzkiej świadomości.

Co ciekawe, w ten sposób naukowcy często wyglądają jak dualiści, bo kiedy już „przedstawili" rzeczy w świadomości, wydają się zrzucać odpowiedzialność na umysł, a psychologowie poznawczy często wychodzą na zombistów (automatystów?), gdyż opisują struktury nieznane neuroanatomom, a ich teorie utrzymują, że cała praca może zostać wykonana bez potrzeby odwoływania się do jakiegokolwiek wewnętrznego obserwatora.

Pozory mylą. Crick i Koch nie są dualistami (nawet jeśli wydaje się, że są kartezjańskimi materialistami), a psychologowie poznawczy nie *zaprzeczyli* istnieniu świadomości (nawet jeśli najczęściej starają się ją ignorować). Co więcej, te wąskie perspektywy nie dyskwalifikują żadnego z tych przedsięwzięć. Neuronaukowcy mają rację, gdy twierdzą, że nie można otrzymać dobrego modelu świadomości, dopóki nie rozwiąże się problemu jej lokalizacji w mózgu, ale kognitywiści (na przykład specjaliści od sztucznej inteligencji i psychologowie poznawczy) mają rację, gdy mówią, że nie można mieć dobrego modelu świadomości, dopóki nie rozwiąże się problemu, jakie funkcje spełnia oraz jak to robi – mechanicznie, bez pomocy umysłu. Jak powiada Philip Johnson-Laird (1983, s. 477): „każda naukowa teoria umysłu musi traktować go jak automat". Ograniczone perspektywy wszystkich tych dyscyplin pokazują nam, że potrzeba kolejnej perspektywy – którą tu kreślimy my – łączącej jak najwięcej mocnych stron każdej z nich.

2. Miniatura jako punkt wyjścia

Moje główne zadanie w tej książce jest natury filozoficznej: pokazać, jak *można by* skonstruować naprawdę wyjaśniającą teorię świadomości z tych części, a nie dostarczyć – i potwierdzić – taką teorię we wszystkich szczegółach. Jednak moja teoria byłaby niepojmowalna (przynajmniej dla mnie), gdyby w ogromnym stopniu nie zapożyczała z wyników empirycznych uzyskanych w różnych dziedzinach. Te wyniki otworzyły (przynajmniej mi) oczy na nowe pomysły. (Szczególnie bogaty zbiór wyników empirycznych i nowych pomysłów na temat świadomości można znaleźć u Marcela i Bisiacha 1988). Mamy wspaniałe czasy na badania nad umysłem. Atmosfera jest gęsta od nowych odkryć, nowych modeli, zaskakujących wyników empirycznych – i mniej więcej tyluż przereklamowanych „dowodów" i przedwczesnych dyskredytacji. Obecnie granica badań nad umysłem jest tak szeroko otwarta, że nie istnieje właściwie żadna powszechnie przyjęta norma wyznaczająca, jakie pytania i metody są właściwe. Przy tak ogromnej ilości słabo uzasadnionych fragmentów teorii i spekulacji warto odłożyć na później żądania dowodu, a zamiast tego przyjrzeć się mniej lub bardziej niezależnym, ale również nierozstrzygającym racjom, które zwykle łącznie mają podbudować poszczególne hipotezy. Powinniśmy jednak trzymać nasz entuzjazm na wodzy. Czasem to, co wydaje się dymem świadczącym o gigantycznym pożarze, jest tak naprawdę jedynie chmurą kurzu pochodzącą z przejeżdżającego wozu z awanturnikami.

Psycholog Bernard Baars w swojej książce *A Cognitive Theory of Consciousness* podsumowuje to, co jego zdaniem jest „rosnącym konsensusem", że świadomość powstaje w wyniku „rozproszonego oddziaływania społeczności specjalistów, wyposażonej w pamięć roboczą, zwaną *globalną przestrzenią roboczą*, której treści mogą zostać udostępnione całemu systemowi" (Baars 1988, s. 42). Jak zauważa, różni teoretycy, mimo ogromnych różnic poglądów, wykształcenia i celów, skłaniają się ku tej wspólnej wizji, jak świadomość musi lokować się w mózgu. Ostrożnie przedstawiam tu wersję tego wyłaniającego się konsensusu, pomijając jedne kwestie, a podkreślając inne – kwestie, które według mnie albo są pomijane, albo

niedoceniane, a okazują się kluczowe w zgłębieniu nadal istniejących tajemnic pojęciowych.

Aby porównać moją teorię do dziesiątek prac, z których wiele zapożyczyła, spójrzmy jeszcze raz na jej miniaturę, krok po kroku zarysowując podobieństwa, wskazując na źródła i niezgodności.

Nie istnieje jeden definitywny „strumień świadomości", gdyż nie ma głównej siedziby czy teatru kartezjańskiego, gdzie „wszystko łączy się w jedno" na użytek centralnego nadawacza sensu. [...]

Wszyscy zgadzają się co do tego, że nie istnieje jedno takie miejsce w mózgu, podobne do Kartezjańskiej szyszynki, jednak wynikające z tego konsekwencje nie zostały rozpoznane i zdarza się, że są w sposób niedopuszczalny pomijane. Na przykład nieroztropne formułowanie „problemu scalania" w obecnych badaniach neuronaukowych często zakłada, że musi istnieć jakaś konkretna, reprezentacyjna przestrzeń w mózgu (mniejsza niż cały mózg), gdzie łączą się wyniki wszelkich rozróżnień – podkład dźwiękowy z filmem, kolory z kształtami, a luki z wypełnieniami. Istnieją pewne ostrożne sformułowania problemu (lub problemów) scalania unikające tego błędu, lecz często brak w nich precyzji.

[...] Zamiast takiego pojedynczego strumienia (choćby i szerokiego) istnieje wiele kanałów, w których specjalistyczne obwody próbują, w równoległych pandemoniach, zajmować się swoimi obowiązkami, tworząc po drodze wielokrotne szkice. Większość tych fragmentarycznych szkiców „narracji" odgrywa krótkotrwałe role, modulując bieżące czynności [...]

W dziedzinie sztucznej inteligencji znaczenie ciągów przypominających narracje od dawna podkreśla Roger Schank; najpierw w pracy o *skryptach* (1977, wraz z Robertem Abelsonem), a potem (1991) w badaniach nad rolą opowiadania historii w rozumieniu. Ze zdecydowanie innych perspektyw, nadal na polu sztucznej inteligencji, Patrick Hayes (1979/2003), Marvin Minsky (1975), John Anderson (1983) i Erik Sandeval (1991) – oraz inni – kładą nacisk na znaczenie struktur danych niebędących jedynie ciągami „migawek" (z towarzyszącym temu problemem ponownej identyfikacji szczegółów w następujących po sobie klatkach), a zamiast tego mają konstrukcję służącą w ten czy inny sposób do bezpośredniego reprezentowania ciągów czasowych i rodzaju ciągów. W filozofii Gareth Evans (1982) kreślił pewne pokrewne idee przed swą przedwczesną śmiercią. W neurobiologii te fragmenty narracyjne są badane jako *scenariusze* i inne ciągi w podejściu zwanym „maszyną Darwina" autorstwa Williama Calvina (1987). Antropologowie od dawna utrzymują, że mity, które każda kultura przekazuje swoim nowym członkom, odgrywają ważną rolę w kształtowaniu ich umysłów (zob. np. Goody 1977, a w kwestii sugerowanego zastosowania w dziedzinie sztucznej inteligencji – Dennett 1991b), jednak nie próbowali stworzyć dla tej idei modelu ani obliczeniowego, ani neuroanatomicznego.

[...] jednak niektóre dostają awans i otrzymują kolejne role funkcjonalne, raz za razem, dzięki działaniu maszyny wirtualnej w mózgu. Szeregowość owej maszyny (jej „von neumannowski" charakter) nie jest cechą zainstalowaną na stałe, a raczej rezultatem następujących po sobie zgrupowań tych specjalistów. [...]

Wielu zauważyło stosunkowo powolne, osobliwe tempo świadomej czynności umysłowej (np. Baars 1988, s. 120) i od dawna czai się sugestia, iż może to być spowodowane tym, że mózg nie był tak naprawdę zaprojektowany – zbudowany sprzętowo – pod kątem tej czynności. Od kilku lat słyszy się o idei, że ludzka świadomość może w związku z tym być działaniem pewnego rodzaju szeregowej maszyny wirtualnej, implementowanej na równoległym mózgowym sprzęcie. Psycholog Stephen Kosslyn zaproponował wersję idei szeregowej maszyny wirtualnej podczas spotkania Society for Philosophy and Psychology we wczesnych latach osiemdziesiątych XX

wieku, a ja zastanawiam się nad różnymi wersjami tego pomysłu mniej więcej od tego samego czasu (np. Dennett 1982b), ale wcześniejszą prezentację właściwie tego samego pomysłu – chociaż bez użycia pojęcia „maszyna wirtualna" – można znaleźć w doniosłej pracy psychologa Paula Rozina *The Evolution of Intelligence and Access to the Cognitive Unconscious* (1976). Inny psycholog, Julian Jaynes, w swoich bezczelnie oryginalnych spekulacjach w *The Origin of Consciousness in the Breakdown of the Bicameral Mind* (1976) podkreślał, że ludzka świadomość jest bardzo świeżym i kulturowym obciążeniem wcześniejszej funkcjonalnej architektury, co jest kwestią rozwiniętą na inne sposoby również przez neuronaukowca Harry'ego Jerisona (1973). Według niego leżąca u podstaw architektura neuronalna to wcale nie *tabula rasa* w momencie narodzin, ale mimo wszystko jest ośrodkiem, w którym budowane są struktury w wyniku interakcji mózgu ze światem zewnętrznym. I to te struktury, bardziej niż te wrodzone, muszą być wskazywane w celu wyjaśnienia funkcjonowania poznawczego.

Podstawowi specjaliści należą do naszego dziedzictwa zwierzęcego. Nie powstali po to, aby wykonywać poszczególne czynności ludzkie, jak czytanie czy pisanie, ale takie jak uchylanie się, unikanie drapieżników, rozpoznawanie twarzy, chwytanie, rzucanie, zbieranie jagód i inne kluczowe umiejętności. [...]

Istnienie tych hord specjalistów potwierdzają różne teorie, lecz sporne są ich wielkość, role czy organizacja. (Allport przedstawia użyteczny, zwięzły przegląd – 1989, s. 643–647). Neuroanatomowie badający mózgi zwierząt, od morskich ślimaków i kałamarnic począwszy, na kotach i małpach skończywszy, wskazali wiele rodzajów wbudowanych obwodów zaprojektowanych wyłącznie do wykonywania konkretnych zadań. Biolodzy mówią o wrodzonych mechanizmach wyzwalających [ang. *Innate Releasing Mechanisms* – IRM] oraz o sztywnych wzorcach ruchowych [ang. *Fixed Action Patterns* – FAP], które mogą się ze sobą połączyć, i w niedawnym liście do mnie neuropsycholożka Lynn Waterhouse trafnie opisała umysły zwierząt jako powstałe z „przeplatających się nawzajem IRM-ów i FAP-ów". Właśnie co do tak problematycznie splecionych zwierzęcych umysłów Rozin (wraz z innymi) zakłada, że są podstawą do ewolucji w umysły o bardziej ogólnym przeznaczeniu, które wykorzystują istniejące wcześniej mechanizmy dla nowych celów. Psycholog percepcji Vilayanur S. Ramachandran (1991) pisze, że „jest to faktyczna zaleta, która przejawia się w wielu systemach: zapewnia ci tolerancję na hałaśliwe obrazy, które spotykasz w rzeczywistym świecie. Moja ulubiona analogia, którą stosuję, aby zilustrować niektóre z tych pomysłów, jest taka, że jest to trochę jak dwóch pijaków; żaden z nich nie może chodzić bez podpierania się, ale opierając się na sobie nawzajem, potrafią jakoś doczłapać do celu".

Neuropsycholog Michael Gazzaniga wskazuje na olbrzymią ilość danych wynikających z neurologicznych deficytów (w tym słynnych, ale często źle opisywanych, pacjentów z rozszczepieniem mózgu), które wspierają pogląd, że umysł to koalicja lub zbitka na wpół niezależnych agencji (Gazzaniga i Ledoux 1978; Gazzaniga 1985); a wychodząc z innych założeń, filozof psychologii Jerry Fodor (1983) twierdzi, że spore części ludzkiego umysłu składają się z *modułów*: wbudowanych, służących konkretnemu celowi, „izolowanych" systemów analizy wejściowej (oraz tworzenia wyjścia – choć o tym nie miał wiele do powiedzenia).

Fodor koncentruje się na modułach, które miałyby być swoiste dla ludzkiego mózgu – szczególnie modułach do nabywania języka i do analizy składniowej zdań – a skoro w dużej mierze pomija kwestię prawdopodobnych przodków w umysłach zwierząt niższych, tworzy nieprawdopodobną wizję ewolucji, która zaprojektowała zupełnie nowe mechanizmy językowe charakterystyczne tylko dla naszego gatunku, z zupełnie nowego materiału. Ten obraz owych modułów jako cudownego podarunku od Matki Natury dla *Hominum sapientum* opiera się na

skrajnie intelektualistycznych poglądach Fodora na temat połączenia modułów z resztą mózgu. Według Fodora nie wykonują pełnych zadań w ekonomii umysłu (jak na przykład kontrola koordynacji wzrokowo-ruchowej przy podnoszeniu przedmiotu), ale nagle zatrzymują się przy wewnętrznej krawędzi, linii w umyśle, której nie mogą przekroczyć. Fodor twierdzi, że istnieje centralna arena racjonalnego „ustalania przekonań", do której moduły służalczo dostarczają swoje dobra, a tam przekazując je do *nie*modularnych („globalnych, izotropowych") procesów.

Moduły Fodora są marzeniem biurokraty: opisy ich obowiązków są sztywne i niezmienne; nie mogą zostać wykorzystane do odegrania nowej roli lub wielu ról; i są „poznawczo nieprzenikalne" – co oznacza, że ich aktywność nie może być przekształcana, ani nawet przerwana, przez zmiany w „globalnych" stanach informacyjnych reszty systemu. Według Fodora wszystkie naprawdę głębokie czynności poznawcze są niemodularne. Myślenie o tym, co zrobić później, rozumowanie o hipotetycznych sytuacjach, twórcze przekształcanie materiałów, zmiana spojrzenia na życie – wszystkie te czynności są wykonywane przez tajemniczą, centralną władzę poznawczą. Co więcej, Fodor twierdzi (z osobliwą satysfakcją), że żadna gałąź kognitywistyki – w tym filozofia – nie ma żadnego pojęcia o tym, jak owa centralna władza funkcjonuje!

Wiele wiemy o przemianach reprezentacji, które służą do uzyskiwania informacji w formie odpowiedniej dla centralnego przetwarzania; nie wiemy właściwie nic o tym, co się dzieje, gdy informacja tam dotrze. Duch został zwabiony w głąb maszyny, ale nie wyegzorcyzmowany. [Fodor 1983, s. 127]

Dając tej centralnej władzy tyle do roboty oraz tyle niemodularnej mocy, z którą może oddziaływać, Fodor zamienił swoje moduły w mało prawdopodobne podmioty, których egzystencja nabiera sensu dopiero w towarzystwie ich Szefa o złowrogiej władzy (Dennett 1984b). Ponieważ jedną z głównych dla Fodora kwestii w zakresie opisywania modułów było skontrastowanie ich skończonej, zrozumiałej, bezmyślnej automatyzacji z nieograniczoną i niewyjaśnialną władzą centrum, teoretycy, którzy w innym razie zaakceptowaliby przynajmniej większość z cech modułów, zwykle traktują tę ideę jako kryptokartezjańską fantazję.

Wielu tych samych teoretyków jest nastawionych neutralnie bądź negatywnie do *agentów* Marvina Minsky'ego, tworzących społeczność umysłu (*The Society of Mind*, 1985). Agenty Minsky'ego to homunkulusy wszelakiej wielkości, od ogromnych specjalistów z talentami tak rozwiniętymi, jak w przypadku modułów Fodora, aż po agentów wielkości memów (polinemów, mikronemów, cenzorów, eliminatorów i wielu innych). Wszystko to wygląda na zbyt łatwe, myślą sceptycy. Gdy pojawia się zadanie, przyjmij grupę agentów na miarę tego zadania, którzy je wykonają – teoretyczny ruch, ze wszystkimi zaletami kradzieży nad prawdziwą pracą, żeby posłużyć się słynnym powiedzonkiem Bertranda Russella.

Homunkulusy – demony, agenty – są walutą w sztucznej inteligencji i bardziej ogólnie w informatyce. Każdy, kto sceptycznie krzywi się na samo wspomnienie homunkulusa, zwyczajnie nie rozumie, jak neutralnym i szeroko stosowanym pojęciem może być. Zakładanie istnienia grupy homunkulusów rzeczywiście byłoby gestem pustym, dokładnie tak jak sądzi sceptyk, gdyby nie fakt, że w teoriach z homunkulusami poważne treści kryją się w tym, jak przyjęte przez teorię homunkulusy wchodzą w interakcje ze sobą, rozwijają się, tworzą koalicje lub hierarchie itd. I tutaj teorie rzeczywiście mogą być bardzo różne. Teorie biurokratyczne, jak widzieliśmy w rozdziale 8, organizują homunkulusy we wcześniej zaprojektowane hierarchie. Nie ma bitwy na poduszki lub nieusłuchanych homunkulusów, a konkurencja pomiędzy nimi jest ściśle regulowana niczym ekstraliga baseballowa. Teorie pandemonium, przeciwnie, zakładają mnóstwo marnotrawstwa, daremnych ruchów, zakłóceń, okresów chaosu oraz obiboków bez ustalonego zakresu obowiązków. Nazywanie jednostek w obu tych różniących się znacznie od

siebie rodzajach teorii „homunkulusami" (lub „demonami" czy „agentami") jest niewiele bardziej treściwe niż nazywanie ich po prostu... „jednostkami". Są po prostu jednostkami ze szczególnie opisanymi kompetencjami i każda teoria, od najbardziej rygorystycznie neuroanatomicznej aż do najbardziej abstrakcyjnie sztucznej, zakłada jednostki tego rodzaju, a następnie teoretyzuje o tym, jak większe funkcje mogą powstawać na mocy organizacji jednostek wykonujących mniej znaczące czynności. Tak naprawdę *wszystkie* odmiany funkcjonalizmu mogą być postrzegane jako funkcjonalizm „homunkularny" o takiej czy innej szczegółowości.

Rozśmieszył mnie niedawno pewien eufemizm, który ostatnio zyskuje uznanie wśród neuronaukowców. Neuroanatomiści poczynili ogromny postęp, sporządzając mapę kory mózgowej, która okazała się niezwykle zorganizowana w specjalistyczne kolumny komunikujących się ze sobą neuronów (neuronaukowiec Vernon Mountcastle, 1978, nazywa je „modułami jednostkowymi"), które następnie są organizowane w większe organizacje, takie jak „mapy retinotopowe" (na których zachowany jest przestrzenny schemat pobudzenia siatkówki oczu), a które z kolei odgrywają role – nadal źle pojmowane – w jeszcze większych organizacjach neuronów. Kiedyś neuronaukowcy dyskutowali o tym, co te różne obszary grup neuronów w korze *sygnalizowały*; myśleli o tych jednostkach jak o homunkulusach, których „zadaniem" zawsze było „wysłanie informacji o konkretnej treści". Niedawne postępy teoretyczne sugerują, że te obszary pełnią o wiele bardziej kompleksowe i różnorodne funkcje, więc za istotnie mylące uważa się teraz mówienie o nich jako o (tylko) sygnalizujących to czy tamto. Jak zatem możemy wyrazić tak mozolnie uzyskane odkrycia dotyczące specyficznych warunków, w których te obszary się uaktywniają? Mówimy, że dla tego obszaru „ważny jest" kolor, podczas gdy dla innego „ważna jest" lokalizacja czy ruch. Jednak to użycie nie jest absurdalnym antropomorfizowaniem bądź też „błędem zakładania homunkulusa" w rodzaju tego, które spotykamy wszędzie na polu sztucznej inteligencji! Oczywiście, że nie. Jest to po prostu mądry sposób, który trzeźwo myślący badacze odnaleźli na potrzeby rozmowy, sugestywnej, ale bez zbytniej szczegółowości, o kompetencjach obszarów nerwowych! Co ujdzie jednemu, nie ujdzie innemu.

Agenty Minsky'ego wyróżniają się głównie dlatego, że w przeciwieństwie do niemal każdej odmiany zakładanych homunkulusów mają historie i genealogie. Ich istnienie nie jest po prostu założone; musiały rozwinąć się z czegoś, co we wcześniejszym istnieniu nie było całkowitą zagadką, a Minsky ma wiele propozycji co do ich możliwego rozwoju. Jeśli nadal irytująco nie ma zdania na temat tego, z jakich neuronów zrobione są agenty oraz gdzie w mózgu się znajdują, to tylko dlatego, że chciał zbadać najogólniejsze wymagania, lecz bez zbytniego upraszczania, związane z rozwojem funkcji. Jak mówi, opisując swoją wcześniejszą teorię „ram" (której potomkiem jest *The Society of Mind*), „gdyby ta teoria była bardziej mglista, zostałaby zignorowana, jednak gdyby została opisana dokładniej, inni naukowcy mogliby ją »przetestować«, zamiast włączyć do niej swoje własne pomysły" (Minsky 1985, s. 259). Niektórzy naukowcy są niewzruszeni na tę obronę. Interesują się jedynie tymi teoriami, które dostarczają sprawdzalnych przewidywań już teraz. Byłaby to dobra, trzeźwa polityka, gdyby nie fakt, że wszystkie dotychczas wymyślone sprawdzalne teorie są, jak można wykazać, błędne i jest głupotą myślenie, że przełomy teoretyczne wymagane do stworzenia *nowych* sprawdzalnych teorii znajdą się nie wiadomo gdzie bez pomysłowej eksploracji w rodzaju tej, w którą zagłębia się Minsky. (Ja oczywiście gram w tę samą grę).

Wracając do teorii w miniaturze:

Często są [specjalistyczne demony] oportunistycznie werbowane do nowych ról, do których mniej więcej pasują ich przyrodzone talenty. Rezultatem nie jest chaos tylko dlatego, że tendencje narzucone tym czynnościom same w sobie są wynikiem konstrukcji. Część tej

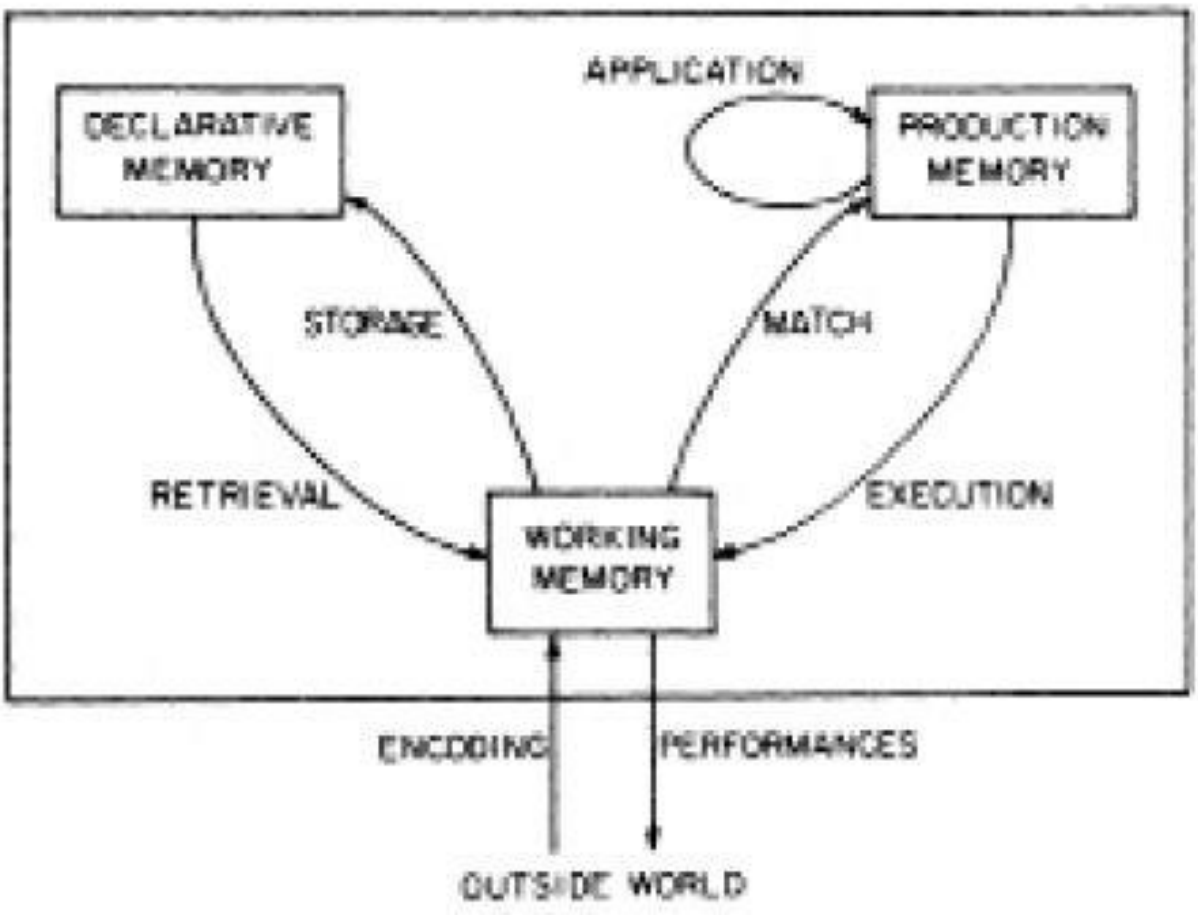
konstrukcji jest wrodzona i dzielimy ją z innymi zwierzętami. Konstrukcja ta jest jednak udoskonalona, a czasem nawet zupełnie zmarginalizowana przez mikronawyki myślowe rozwijające się w indywidualnych, częściowo specyficznych rezultatach autoeksploracji, a częściowo w gotowych wytworach kultury. Tysiące memów, zrodzonych głównie przez język, ale również przez bezsłowne „obrazy" i inne struktury danych, zamieszkują indywidualny mózg, kształtując jego tendencje i w ten sposób przemieniając go w umysł.

W tej części mojej teorii celowo nie dawałem odpowiedzi na wiele ważnych pytań: Jak właściwie te homunkulusy komunikują się ze sobą, aby coś osiągnąć? Jakie są bazowe operacje przetwarzania informacji i jakie mamy racje, aby sądzić, że mogą „działać"? Według tego szkicu kolejność zdarzeń jest wyznaczona (w sposób, który jedynie mgliście sugerowałem) przez „nawyki", a oprócz pewnych negatywnych twierdzeń z rozdziału 5 dotyczących tego, co się *nie* dzieje, nie wyraziłem się jeszcze jasno na temat struktury procesów, dzięki którym elementy wielokrotnych szkiców zostają zachowane, niektóre z nich zaś w końcu tworzą heterofenomenologię w rezultacie takiego czy innego sondowania. Aby zobaczyć, do czego sprowadza się to pytanie oraz jakie mogłyby być konkurencyjne odpowiedzi, powinniśmy na chwilę spojrzeć na pewne lepiej zarysowane modele myślenia sekwencyjnego.

3. A co potem?

W rozdziale 7 widzieliśmy, że architektura von neumannowska jest kwintesencją szeregowych procesów celowych obliczeń. Turing i von Neumann wyodrębnili jeden szczególny rodzaj nurtu mogącego przepływać przez strumień świadomości, a następnie radykalnie wyidealizowali go, chcąc go zmechanizować. Jest tam owo ciesząca się złą sławą wąskie gardło von Neumanna, jeden rejestr na wyniki i jeden rejestr na instrukcje. Programy to po prostu uporządkowane listy instrukcji z małego zbioru poleceń elementarnych, które są wbudowane w maszynę. Ustalony proces, cykl „pobierz-wykonaj", pobiera instrukcje z kolejki w pamięci, jedną po drugiej, zawsze biorąc kolejną instrukcję z listy, chyba że wcześniejsza instrukcja była skokiem do innej części listy.

Gdy twórcy modeli sztucznej inteligencji zaczęli implementować bardziej realistyczne modele operacji poznawczych na takiej podstawie, zrewidowali to wszystko. Rozszerzyli skandalicznie wąskie gardło von Neumanna, zamieniając je w zwięzłą „przestrzeń roboczą" czy „pamięć roboczą". Zaprojektowali również bardziej zaawansowane operacje, które stały się elementarnymi poleceniami psychologicznymi, oraz zastąpili sztywny cykl „pobierz-wykonaj" maszyny von Neumanna elastyczniejszymi sposobami wywoływania i wykonywania instrukcji. Przestrzeń robocza stała się w niektórych przypadkach „tablicą" (Reddy i inni 1973; Hayes-Roth 1985), gdzie różne demony mogą zostawiać informacje dla innych demonów, co z kolei wywołuje kolejną falę pisania i czytania. Architektura von Neumanna ze swoimi sztywnymi cyklami instrukcji nadal znajdowała się w tle jako podstawa implementacji, ale nie odgrywała roli w tym modelu. To, co następnie działo się w modelu, zależało od wyników konkurujących ze sobą fal zapisów i odczytów komunikatów z tablicy. Spadkobiercami idei maszyny von Neumanna są różne *systemy produkcji* (Newell 1973), będące podstawą takich modeli, jak ACT* Johna Andersona (1983) oraz Soar Rosenblooma, Lairda i Newella (1987) (zob. również Newell 1990).



Ryc. 9.1

Wewnętrzna architektura systemów produkcji jest widoczna na ilustracji ukazującej ACT* (Ryc. 9.1).

To w pamięci roboczej dzieje się wszystko. Wszelkie podstawowe działania nazywane są *produkcjami*. Produkcje są zwykle po prostu mechanizmami rozpoznawania wzorców, które uruchamiają się, ilekroć wykryją *swój* określony wzorzec. Innymi słowy, są czekającymi w pogotowiu operatorami JEŚLI-TO, monitorującymi bieżącą zawartość pamięci roboczej i czekającymi na moment, w którym zostanie spełniony ich warunek JEŚLI, a wówczas mogą zrealizować wymaganą operację (w klasycznym systemie produkcji jest nią umieszczenie nowej informacji w pamięci roboczej do dalszego użytku produkcji).

Wszystkie komputery mają elementarne instrukcje JEŚLI-TO, „organy zmysłów" umożliwiające reagowanie na dane przychodzące lub wydobywane z pamięci. Ta umiejętność *skoków warunkowych* jest podstawowym czynnikiem mocy komputerów – bez względu na to, na jakiej architekturze się opierają. Pierwotne instrukcje JEŚLI-TO były prostymi, jednoznacznymi instrukcjami dotyczącymi stanów maszyny Turniga: JEŚLI widzisz zero, TO zamień je na jedynkę, przesuń się o jedno pole w lewo i przejdź w stan *n*. Porównajmy tak proste polecenia z takimi, które możesz dać dobrze wyszkolonemu i doświadczonemu ludzkiemu strażnikowi: JEŚLI zobaczysz coś, co wygląda na nieznany obiekt ORAZ następujące potem badanie nie przyniesie żadnych rezultatów LUB pozostawia jakieś wątpliwości, TO należy wszcząć alarm. Czy możemy stworzyć tak zaawansowany system monitoringu z prostych JEŚLI-TO? *Produkcje*

są pośrednimi sensorami, za pomocą których można wybudować bardziej kompleksowe organy zmysłów, a następnie całą architekturę poznawczą. Produkcje mogą przyjmować złożone i rozmyte warunki JEŚLI; wzorce, które „rozpoznają", nie muszą być tak proste, jak kody kreskowe rozpoznawane przez kasy, ale przypominać raczej wzorce identyfikowane przez strażnika (zob. rozważania na ten temat w Anderson 1983, s. 35–44). A w przeciwieństwie do instrukcji JEŚLI-TO maszyny Turinga, która jest zawsze w jednym stanie jednocześnie (badając jedno JEŚLI-TO z całego zestawu, a następnie przechodząc do kolejnych danych), instrukcje JEŚLI-TO w systemie produkcyjnym czekają wszystkie, w (symulowanej) równoległości i w każdym „momencie" więcej niż jedna produkcja może mieć spełniony warunek i być gotowa do akcji.

Tu zaczyna się robić ciekawie: W jaki sposób taki system radzi sobie z *rozwiązywaniem konfliktów*? Gdy więcej niż jedna produkcja zostaje spełniona, zawsze istnieje możliwość, że dwie (lub więcej) będą miały niezgodne wskazania. Systemy równoległe mogą tolerować duże ilości niewspółgrających ze sobą celów, jednak w systemie, który ma działać skutecznie w świecie, nie wszystko może się wydarzyć naraz; czasem z czegoś trzeba zrezygnować. Kluczową sprawą odróżniającą różne modele jest ich radzenie sobie z konfliktem. Tak naprawdę, skoro większość, jeśli nie wszystkie, psychologicznie i biologicznie interesujących szczegółów widać w postaci różnic na tym poziomie, to najlepiej postrzegać architekturę systemów produkcji jako materiał, z którego buduje się modele. Jednak wszystkie systemy produkcji mają kilka podstawowych założeń, łączących je z naszą teorią w miniaturze: mają przestrzeń roboczą, w której dzieje się wszystko, gdzie wiele produkcji (= demonów) może próbować uzyskać swój cel jednocześnie, i ma mniej więcej bierną pamięć, w której przechowywane są wrodzone i nagromadzone informacje. Ponieważ nie wszystko, co jest „znane" takiemu systemowi, jest dostępne w tej przestrzeni roboczej w tym samym czasie, problem Platona polegający na przywołaniu odpowiedniego ptaka w odpowiednim momencie staje się poważnym zadaniem logistycznym. A z naszego obecnego punktu widzenia najważniejsze jest to, że teoretycy *stworzyli* mechanizmy mogące posłużyć do odpowiedzi na trudne pytanie: *co potem*?

Na przykład w ACT* istnieje pięć zasad rozwiązywania konfliktu:

(1) *Stopień dopasowania*: Gdy warunek JEŚLI jednej produkcji jest pod jakimś względem lepiej dopasowany niż innej, staje się priorytetem.

(2) *Siła produkcji*: Produkcje, które w ostatnim czasie były udane, mają powiązaną ze sobą większą „siłę", co daje im priorytet nad produkcjami o mniejszej sile.

(3) *Oporność danych*: Ta sama produkcja nie może zostać dopasowana do tych samych danych więcej niż raz (aby uniknąć niekończących się zapętleń i podobnej, choć mniej drastycznej, rutynowości).

(4) *Specyficzność*: Gdy dwie produkcje pasują do tych samych danych, wygrywa ta z bardziej skonkretyzowanymi warunkami JEŚLI.

(5) *Dominacja celów*: Wśród elementów przechowywanych w pamięci roboczej są cele. Może istnieć jeden aktywny cel w danym momencie w pamięci roboczej ACT* i każda produkcja, której wynik pasuje do aktywnego celu, otrzymuje priorytet.

Wszystko to są prawdopodobne zasady rozwiązywania konfliktów, mające zarówno sens

psychologiczny, jak i teleologiczny (szczegółowe uzasadnienia podaje Anderson 1983, rozdz. 4). Być może jednak mają zbyt duży sens. Innymi słowy, Anderson sam mądrze zaprojektował system rozwiązywania konfliktów w ACT*, wykorzystując swoją wiedzę o specyficznych rodzajach problemów pojawiających się w warunkach rozwiązywania konfliktu oraz o efektywnych sposobach radzenia sobie z nimi. Dosłownie wbudował tę zaawansowaną wiedzę w system, wrodzony dar od ewolucji.

Interesującym kontrastem jest architektura Soar Rosenblooma, Lairda i Newella (1987). Ona również, jak każda architektura równoległa, napotyka *impasy* – sytuacje, w których zachodzi potrzeba rozwiązania konfliktu, gdyż pojawiają się przeciwne produkcje lub brakuje produkcji – jednak traktuje je jako dobrodziejstwo, a nie problem. Impasy są podstawowymi szansami w systemie. Konflikty nie są rozwiązywane automatycznie przez wcześniej proroczo ustalone zbiory reguł rozwiązywania konfliktów (autorytatywne homunkulusy-policjanci z drogówki ustawione wcześniej na miejscu), ale są rozwiązywane *nie*automatycznie. Impas tworzy nową „przestrzeń problemową" (rodzaj aktualnej przestrzeni roboczej), w której problemem do rozwiązania jest właśnie ten impas. Może to wytworzyć kolejną meta-metadrogową przestrzeń problemową i tak dalej – *potencjalnie* w nieskończoność. Jednak w praktyce (przynajmniej w domenach modelowanych do dziś) po powstaniu kilku kolejnych przestrzeni problemowych najwyższy problem znajduje rozwiązanie, które szybko rozwiązuje kwestię znajdującą się niżej i tak dalej, eliminując złowrogi rozrost przestrzeni po niebanalnych badaniach w logicznej przestrzeni możliwości. Poza tym efektem dla systemu jest „przyłączenie" nowo zdobytych odkryć do nowych produkcji i w momencie, gdy w przyszłości pojawi się podobny problem, będzie już dostępna nowo powstała produkcja, szybko rozwiązująca banalny problem, wcześniej już rozwiązany.

Pokrótce przedstawiam te szczegóły nie po to, aby przekonywać do ostatecznej wyższości systemu Soar nad ACT*, a jedynie po to, by pokazać pewien pomysł, który można wykorzystywać, w sposób odpowiedzialny, w modelach budowanych na podobnej zasadzie. Mam przeczucie, że z różnych powodów, które nie są w tej chwili dla nas istotne, wewnętrzny nośnik systemów produkcji jest *nadal* zbyt wyidealizowany i uproszczony w swoich ograniczeniach, lecz droga od maszyny von Neumanna do systemów produkcji wskazuje na dalsze architektury, mające jeszcze więcej inteligencji, a najlepszym sposobem eksploracji ich mocnych punktów i ograniczeń jest ich zbudowanie i przetestowanie. Tak można przekuć impresjonistyczne i niejasne elementy teorii takich jak moja na prawdziwe, szczegółowe modele ze szczegółami, które mogą zostać przetestowane empirycznie.

Gdy przyjrzeć się różnym twierdzeniom dotyczącym świadomości, które wysuwałem przez ostatnie cztery rozdziały, i zacząć próbę zestawienia ich z takimi modelami systemów poznawczych, pojawia się mnóstwo pytań, lecz nie będę próbował tu na nie odpowiedzieć. Skoro zostawiam wszystkie te kwestie otwarte, mój szkic pozostaje właśnie szkicem, który mógłby swobodnie pasować do całej rodziny istotnie różniących się od siebie teorii. W tej sytuacji nie muszę posuwać się dalej, gdyż filozoficzne problemy świadomości dotyczą tego, czy *którakolwiek* z tych teorii może wyjaśnić świadomość, więc przedwczesne byłoby wzbudzać nadzieję na prawdziwość zbyt szczegółowej wersji, która może się okazać istotnie wadliwa. (W załączniku B dam jednak trochę empirycznych szczegółów, interesujących dla tych, którzy chcą, aby teoria miała od początku sprawdzalne założenia).

Nie tylko teorie filozofów muszą na tym poziomie zdać test modelowania; to samo dotyczy teorii neuronaukowych. Na przykład złożona teoria Geralda Edelmana (1989) obwodów „współbieżnych" w mózgu zawiera wiele twierdzeń dotyczących tego, jak takie obwody realizują rozróżnienia, tworzą struktury pamięciowe, koordynują sekwencje kroków prowadzących do

rozwiązania problemu i ogólnie kierują działaniami ludzkiego mózgu, lecz pomimo bogactwa detali neuroanatomicznych oraz entuzjastycznych i często wiarygodnych zapewnień Edelmana nie będziemy wiedzieli, co jego współbieżne elementy mogą robić – nie będziemy wiedzieli, że współbieżne elementy są *prawidłowym* sposobem pojmowania funkcjonalnej neuroanatomii – dopóki nie zostaną one włączone do całej architektury kognitywnej na poziomie najmniejszych struktur, jak to jest w przypadku ACT* czy Soar, i nie będą przetestowane[81].

Na bardziej szczegółowym poziomie modelowania pozostaje nierozwiązana kwestia, jak produkcje (jeśli tak nazwiemy demony rozpoznające wzorce) same w sobie są zaimplementowane w mózgu. Baars (1988) nazywa swoich specjalistów „cegiełkami" i argumentuje za odłożeniem pytania, skąd się biorą cegiełki, na później lub przekazanie tego pytania innej dyscyplinie, lecz – jak zauważyło wielu – kuszące jest założenie, że sami specjaliści na kilku poziomach agregacji powinni być modelowani jako elementy jakiegoś rodzaju tkaniny *koneksjonistycznej*.

Koneksjonizm (rozproszone przetwarzanie równoległe, ang. *parallel distributed processing* [PDP]) jest raczej świeżym pomysłem w dziedzinie sztucznej inteligencji, mającym zbliżyć modelowanie poznawcze do modelowania neuronowego, gdyż elementy, które są *jego* cegiełkami, to węzły w równoległych sieciach połączonych ze sobą tak, że *przypominają* sieci neuronowe w mózgu. Porównywanie koneksjonistycznej sztucznej inteligencji (AI) do „starej, dobrej AI" (Haugeland 1985) oraz do różnych projektów modelowania w dziedzinie neuronauki stało się wręcz osobnym działem w dyskusjach akademickich (zob. np. Graubard 1988; Bechtel i Abrahamson 1991; Ramsey, Stich i Rumelhart 1991). Nie jest to zaskoczeniem, gdyż koneksjonizm przeciera pierwsze w miarę wiarygodne szlaki prowadzące do unifikacji ogromnego, niezbadanego obszaru leżącego między nauką o umyśle a nauką o mózgu. Jednak niemal żaden ze sporów wokół „odpowiedniego traktowania koneksjonizmu" (Smolensky 1988) nie ma wpływu na nasz projekt. *Oczywiście*, że będzie musiał powstać poziom (lub poziomy) teorii tak samo dokładny jak modele koneksjonistyczne i że będzie musiał znajdować się między bardziej oczywistymi nauroanatomicznymi poziomami teorii a bardziej oczywistymi psychologicznymi czy poznawczymi poziomami teorii. Pytanie brzmi, dokładnie które pomysły koneksjonistyczne będą należeć do rozwiązania, a które odpadną po drodze. Dopóki nie jest to wyjaśnione, teoretycy zwykle wykorzystują spory wokół koneksjonizmu jako okazję do głoszenia swoich ulubionych sloganów, a choć chcę, jak każdy, stanąć po którejś stronie tej debaty (Dennett 1987b, 1988b, 1989, 1991b, 1991c, 1991d), to tutaj ugryzę się w język i będę kontynuować nasze najważniejsze zadanie, którym jest zobaczenie, jak teoria *świadomości* może się z tego wyłonić, gdy w którymś momencie jednak skończy się całe zamieszanie.

Spójrzmy, co stało się podczas przejścia od architektury von Neumanna do architektur tak wirtualnych, jak systemy produkcji i (na bardziej szczegółowym poziomie) systemy koneksjonistyczne. Nastąpiło coś, co można by nazwać zmianą równowagi władz. Sztywne, wcześniej zaprojektowane programy, biegnące wzdłuż torów wraz z nielicznymi skokami zależnymi od danych, zostały zastąpione elastycznymi – zmiennymi – systemami, których późniejsze zachowanie jest w większej mierze wynikiem złożonych interakcji między tym, co system chwilowo napotyka, a tym, co napotkał w przeszłości. Jak ujęli to Newell, Rosenbloom i Laird (1989, s. 119): „Tak oto dla standardowego komputera pytaniem jest to, jak przerwać pracę, a kwestią dla Soar czy ACT* (i prawdopodobnie dla ludzkiego postrzegania) jest, jak utrzymać koncentrację".

Biorąc pod uwagę to, ile napisano o tej teoretycznej kwestii, warto podkreślić, że jest to zmiana *równowagi* władz, a nie zmiana na jakiś „jakościowo inny" tryb działania. W sercu najbardziej niestabilnego systemu rozpoznawania wzorców („koneksjonistycznego" lub nie) leży

silnik von Neumanna, terkoczący i obliczający to, co obliczalne. Od momentu narodzin komputerów krytycy sztucznej inteligencji nie mogli darować sobie elaboratów na temat sztywności, mechaniczności, *zaprogramowania* komputerów, a jej obrońcy powtarzali, że kwestią złożoności jest stworzenie na komputerach niesztywnych, rozmytych, holistycznych, organicznych systemów. Wraz z rozwojem sztucznej inteligencji pojawiły się właśnie takie systemy, więc teraz krytycy muszą zdecydować, w tę lub we w tę. Czy powinni na przykład zadeklarować, że systemy koneksjonistyczne są ich zdaniem budulcem umysłów, czy może powinni podnieść stawkę i stwierdzić, że nawet system koneksjonistyczny nie jest dla nich wystarczająco „holistyczny", „intuicyjny" czy... (wpisz swój ulubiony slogan). Dwaj najbardziej znani krytycy sztucznej inteligencji, filozofowie Hubert Dreyfus i John Searle z Uniwersytetu Kalifornijskiego w Berkeley, nie są w tej kwestii zgodni; Dreyfus ślubował wierność koneksjonizmowi (Dreyfus i Dreyfus 1988), podczas gdy Searle podniósł stawkę, twierdząc, że żaden koneksjonistyczny komputer nie jest w stanie cechować się *prawdziwą* umysłowością (Searle 1990a, 1990b).

Sceptycy, którzy obstają za swoim „dla zasady", mogą się wycofywać, ale ogromne problemy nadal trapią zwolenników unifikacji. Moim zdaniem największy jest ten, który ma bezpośredni związek z naszą teorią świadomości. Konsensus w kognitywistyce, który mógłby zostać zilustrowany dziesiątkami diagramów, takich jak na rycinie 9.1, jest taki, że *tam* mamy pamięć długotrwałą (klatka z ptakami Platona), a *tutaj* mamy przestrzeń roboczą, czyli pamięć roboczą, gdzie faktycznie odbywa się myślenie[82]. A jednak nie ma dwóch miejsc w mózgu, które miałyby być siedzibą tych osobnych obiektów. Jedynym miejscem w mózgu, które jest możliwą lokalizacją którejś z tych osobnych funkcji, pozostaje cała kora – nie dwa miejsca obok siebie, lecz jedna duża przestrzeń. Jak mówi Baars, podsumowując tworzący się konsensus, istnieje *globalna* powierzchnia robocza. Jest globalna nie tylko w sensie funkcjonalnym (mówiąc dosadnie: jest to „miejsce", gdzie właściwie wszystko może być w kontakcie z właściwie wszystkim innym), ale również w sensie anatomicznym (jest rozproszona w całej korze i bez wątpienia obejmuje też inne obszary mózgu). Oznacza to więc, że powierzchnia robocza musi wykorzystać te same szlaki neuronowe i sieci, najwyraźniej odgrywające ważną rolę w kwestii pamięci długotrwałej: „pamięć" konstrukcji zmienia się pod wpływem indywidualnej eksploracji.

Załóżmy, że umiesz piec chleb kukurydziany albo wiesz, co znaczy „fenotypowy". W jakiś sposób kora musi być nośnikiem, w którym stabilne schematy połączeń mogą raczej na stałe połączyć te poprawki we wrodzonej konstrukcji mózgu. Załóżmy, iż nagle ci się przypomniało, że masz wizytę u dentysty i znika cała przyjemność, którą czerpiesz ze słuchania muzyki. W jakiś sposób kora musi być nośnikiem, w którym niestabilne schematy połączeń mogą szybko zmienić te krótkotrwałe zawartości całej „przestrzeni" – oczywiście bez jednoczesnego wymazywania pamięci długotrwałej. Jak te dwa zupełnie różne rodzaje reprezentacji mogą współistnieć w tym samym nośniku w tym samym czasie? W modelach czysto poznawczych te zadania mogą być przechowywane w osobnych pudełkach na diagramie, ale gdy musimy je nałożyć na jedną tkaninę tkanki neuronalnej, prosty problem pakowania jest najmniejszym zmartwieniem.

Można założyć, że dwa funkcjonalnie różne systemy sieciowe wzajemnie się przenikają (tak jak system telefoniczny i układ autostrad obejmują cały kontynent) – jednak nie jest to problem. Głębsza kwestia leży zaraz pod powierzchnią – w założeniu, które już przedstawiliśmy. Założyliśmy, że indywidualne, specjalistyczne demony w jakiś sposób wciągają inne w większe struktury. Gdyby była to po prostu kwestia zawołania tych nowych elementów, aby wykorzystać ich *specjalistyczne* talenty do wspólnego celu, mielibyśmy już modele takich procesów – takie jak ACT*, Soar i globalna przestrzeń robocza Baarsa – których szczegóły byłyby w różnym

stopniu przekonujące. Ale co jeśli ci specjaliści są również czasem wciągani *jako znający się na wszystkim*, aby przyczyniać się do realizacji funkcji, w których wyróżniające ich specjalistyczne talenty nie odgrywają żadnej roli? Jest to z różnych powodów kuszący pomysł (zob. np. Kinsbourne i Hicks 1978), choć o ile mi wiadomo, nie mamy jeszcze żadnych obliczeniowych modeli tego, jak mogłyby działać takie elementy o podwójnych funkcjach.

Oto w czym tkwi problem: powszechnie uważa się, że specjaliści w mózgu muszą w jakiś sposób uzyskiwać swoją tożsamość funkcjonalną od swojej aktualnej pozycji w sieci mniej więcej stałych połączeń. Na przykład wydaje się, że jedyny rodzaj faktów, które mogłyby wyjaśnić „odpowiadanie" pewnego szczególnego szlaku neuronowego na barwę, byłyby fakty dotyczące ich specyficznych połączeń, nawet pośrednich, z czopkami na siatkówce oka, które są maksymalnie czułe na różne częstotliwości światła. Gdy taka tożsamość funkcjonalna zostaje ustalona, połączenia te mogą być odcięte (jak dzieje się to w przypadku oślepnięć w wieku dorosłym) bez (całkowitej) straty mocy specjalisty do reprezentowania barw (czy jakiegoś innego „odpowiadania na" nie), ale bez takich przyczynowych połączeń powstałych na początku trudno jest zobaczyć, co mogłoby dać specjaliście rolę swoistą dla tej treści[83]. Wydaje się, że kora (w dużej mierze) składa się z elementów, których mniej lub bardziej stałe moce reprezentacyjne są wynikiem ich funkcyjnej lokalizacji w całej sieci. Reprezentują w sposób, w jaki członkowie Izby Reprezentantów reprezentują regiony: przez przekazywanie informacji ze źródeł, z którymi są bezpośrednio połączeni (na przykład większość ich rozmów na liniach telefonicznych do ich regionów może zostać wywiedziona od ich waszyngtońskich biur). A teraz wyobraźmy sobie członków Izby Reprezentantów siedzących na trybunach na stadionie i reprezentujących ważny slogan „Prędkość zabija!", to znaczy trzymających w górze duże, kolorowe kartki, z których układają się gigantyczne litery tworzące slogan i widoczne z przeciwnej strony stadionu. Żywe piksele, dla których ich relacje ze swoim okręgiem wyborczym nie odgrywają żadnej roli w powstałej reprezentacji grupowej. Pewne modele narastania odpowiedzi korowej silnie sugerują konieczność istnienia *czegoś takiego jak* ta drugorzędna rola reprezentacyjna. Na przykład kuszące jest założenie, że treść informacyjna dotycząca konkretnej kwestii może się pojawić na pewnym specjalistycznym szlaku i wówczas, w jakiś sposób, zostać rozpropagowana w obszarach korowych, wykorzystując niestałość w tych obszarach, bez angażowania się w specjalistyczną semantykę modułów tam rezydujących. Załóżmy na przykład, że nagła zmiana następuje w lewej górnej ćwiartce pola widzenia jakiejś osoby. Jak można się spodziewać, pobudzenie mózgu może być widoczne jako pojawiające się najpierw w częściach kory wzrokowej reprezentujących (jak w przypadku Izby Reprezentantów) różne cechy zdarzeń w lewej, górnej ćwiartce pola widzenia, ale te miejsca natychmiast stają się źródłem rozprzestrzeniającej się aktywacji, obejmującej elementy kory z innymi okręgami wyborczymi. Jeśli to rozprzestrzenienie się pobudzenia w innych obszarach kory nie jest jedynie przeciekiem czy szumem, jeśli odgrywa jakąś kluczową rolę w rozwinięciu lub umożliwieniu redakcji szkicu fragmentu narracji, te dołączone elementy odgrywają inną rolę od tej, gdy są związane z określonym źródłem[84].

Nie jest zaskakujące, że nadal nie mamy dobrych modeli takiej wielofunkcyjności (jedyne wiarygodne spekulacje, z którymi się zetknąłem, zostały przedstawione przez Minsky'ego w *The Society of Mind*). Jak zauważyliśmy w rozdziale 7, ludzcy inżynierowie, pozbawieni możliwości przewidzenia wszystkiego, szkolą się, aby projektować systemy, w których każdy element odgrywa jedną rolę, ostrożnie odizolowany od wpływów z zewnątrz, aby zminimalizować zniszczenia spowodowane przez nieprzewidziane skutki uboczne. Natomiast Matka Natura nie przejmuje się przewidywaniem skutków niepożądanych, więc może je wykorzystać, gdy się przypadkiem pojawią – raz od wielkiego dzwonu. Prawdopodobnie trudność z badaniem

funkcjonalnej dekompozycji kory mózgowej, która dotychczas umyka neuronaukowcom, wynika z faktu, że z natury nie potrafią rozważać hipotezy przypisującej wielu ról dostępnym elementom. Niektórzy romantycy – filozof Owen Flanagan (1991) nazywa ich „neomisterianami" – głoszą istnienie nieprzekraczalnej bariery, uniemożliwiającej mózgowi zrozumienie swojej własnej organizacji (Nagel 1986; McGinn 1989/2008). Ja nie głoszę nic takiego, ale zakładam, że piekielnie trudne – lecz nie niemożliwe – jest zrozumienie, jak działa mózg, częściowo dlatego, że został on zaprojektowany przez proces mogący korzystać z wielu nakładających się funkcji, co jest systematycznie trudne do zidentyfikowania z perspektywy inżynierii odwrotnej.

Są to problemy prowadzące do życzeniowego markowania argumentacji. Niektórych od razu kusi zrezygnowanie z pomysłu takiej dwojakości specyficzność/ogólność – nie dlatego, że mogą udowodnić jej nieprawdziwość, ale dlatego, że nie potrafią sobie wyobrazić, jak stworzyć jej model, a w związku z tym dosyć rozsądnie mają nadzieję, że nigdy nie będą musieli się tym zajmować. Gdy jednak pojawia się taka możliwość, przynajmniej teoretycy muszą szukać nowych wskazówek. Neurofizjologowie (wstępnie) zidentyfikowali mechanizmy w neuronach takich jak receptory NMDA (*N*-metylo-D-asparaginowe) oraz synapsy Malsburga (1985), które mogłyby wiarygodnie odgrywać role szybkich modulatorów połączeń między komórkami. Takie bramki mogłyby pozwolić na sprawne tworzenie tymczasowych „zespołów" mogących być nakładanymi na sieci bez potrzeby jakiejkolwiek zmiany długotrwałych sił synaptycznych, o których zakłada się, że są klejem łączącym ze sobą trwałe zespoły pamięci długotrwałej. (Nowe spekulacje dotyczące tych kwestii znajdziesz we Flohr 1990).

Na większą skalę nauroanatomiści uzupełniają mapę połączeń w mózgu, pokazując nie tylko, które obszary są aktywne w danych warunkach, ale także, co robią. Stawiane są hipotezy zakładające odgrywanie kluczowej dla świadomości roli przez różne obszary. *Układ siatkowaty* w śródmózgowiu i *wzgórze* ponad nim od dawna są znane jako mające zasadniczą rolę w pobudzaniu mózgu – na przykład po śnie lub w odpowiedzi na nowość czy wypadek – a obecnie te ścieżki są lepiej opisane, więc można sformułować i przetestować szczegółowe hipotezy. Na przykład Crick (1984) proponuje, że rozgałęzienia wychodzące ze wzgórza do wszystkich obszarów kory nadają mu rolę „reflektora", w sposób różnorodny pobudzającego lub uwydatniającego konkretne obszary specjalistyczne, wciągając je do realizacji bieżących celów[85]. Baars (1988) rozwinął podobny pomysł: ERTAS, czyli przedłużony system aktywacyjny układu siatkowego i wzgórza (ang. *Extended Reticular-Thalamic Activating System*). Nie byłoby trudne włączenie takiej hipotezy do naszego anatomicznie neutralnego ujęcia współzawodnictwa między koalicjami specjalistów, zakładając, że nie poddamy się kuszącemu obrazowi Szefa we wzgórzu, *rozumiejącego* bieżące wydarzenia, kierowane przez różne części mózgu, z którymi pozostaje „w kontakcie".

Podobnie o *płacie czołowym* kory mózgowej, części mózgu w sposób najwyraźniej powiększonej u *homo sapiens*, wiemy, że jest zaangażowany w długoterminową kontrolę oraz w planowanie i kolejność zachowań. Uszkodzenia różnych obszarów płata czołowego zwykle powodują przeciwstawne symptomy, jak bycie rozkojarzonym i zbytnia koncentracja powodująca trudności w wyjściu poza rutynę oraz impulsywność czy też nieumiejętność poddawania się czynnościom, które zapewniają późniejszą gratyfikację. Kuszące jest więc zainstalowanie Szefa w płacie czołowym i kilka modeli poszło właśnie w tym kierunku. Wyjątkowo zaawansowany model to nadrzędny system uwagi (ang. *Supervisory Attentional System*) Shallice'a (1985), który został ulokowany w korze przedczołowej i któremu nadano szczególną odpowiedzialność za rozwiązywanie konfliktów w przypadku, gdy drugorzędna biurokracja nie daje sobie rady. Po raz kolejny odnalezienie anatomicznej lokalizacji procesów kluczowych dla kontroli tego, co się wydarza, jest jedną kwestią, a inną jest umieszczenie tam

Szefa; każdy, kto poluje na przedni ekran, na którym Szef śledzi kontrolowane przez siebie projekty, szuka wiatru w polu (Fuster 1981; Calvin 1989a).

Gdy już jednak wyrzekniemy się tych kuszących pomysłów, musimy inaczej pojmować udział omawianych obszarów, lecz nadal nie ma tu wielu pomysłów, pomimo ostatnich postępów. Nie chodzi o to, że nie mamy pojęcia, czym jest ta maszyneria; problem jest znacznie bardziej kwestią braku obliczeniowego modelu tego, co wykonuje ta maszyneria i w jaki sposób. Nadal operujemy tu metaforami i odrzucamy niemożliwe, ale nie bójmy się tego etapu; musimy go przebyć, aby dotrzeć do jaśniej sprecyzowanych modeli.

4. Siła maszyny joyce'owskiej

Według naszego szkicu istnieje rywalizacja między wieloma równoczesnymi zdarzeniami mózgowymi, a wybrany ich podzbiór „wygrywa". Oznacza to, że potrafią wywoływać różnego rodzaju stałe efekty. Niektóre z nich, łącząc siły z demonami językowymi, mają wpływ na wypowiedzi, zarówno te wymawiane na głos do innych, jak i te ciche (oraz głośne) wypowiadane do samego siebie. Niektóre przekazują swoją treść innym formom dalszej autostymulacji, jak rysowanie sobie planów. Reszta wymiera niemal natychmiast, pozostawiając jedynie delikatne ślady – poszlaki – tego, że w ogóle kiedyś istniały. Równie dobrze można zechcieć zapytać, co ma na celu zwycięstwo pewnych treści otrzymujących w ten sposób dostęp do zaklętego kręgu – i cóż w nim jest takiego magicznego. Świadomość jest ponoć czymś niezwykle wyjątkowym. Cóż jest wyjątkowego w przepchnięciu do kolejnej rundy w takim kręgu autostymulacji? Jak to pomaga? Czy niemal magiczne moce mają wpływ na zdarzenia zachodzące w takich mechanizmach?

Unikam tezy, że pewien rodzaj wygranej w tej wymuszonej rywalizacji wiąże się z uzyskaniem świadomości. Tak naprawdę uważam, że nie można zasadnie zakreślić granicy między zdarzeniami, które zdecydowanie znajdują się „w" świadomości, a tymi, które na zawsze pozostają „wewnątrz" lub „poniżej" świadomości. (Dalsze argumenty wspierające to stanowisko znajdziesz w Allport 1988). Niemniej jednak, jeśli moja teoria maszyny Joyce'owskiej ma w ogóle uchylić rąbka tajemnicy świadomości, lepiej, żeby pokazała coś szczególnego, związanego z niektórymi, jeśli nie wszystkimi, czynnościami tej maszyny, gdyż nie można zaprzeczać, że świadomość jest intuicyjnie czymś wyjątkowym.

Trudno odpowiedzieć na te znane pytania, nie popadając w pułapkę myślenia, że najpierw należy wyjaśnić, *po co* jest świadomość, aby następnie móc zapytać, czy zaproponowane mechanizmy mogłyby wykonać *tę* funkcję – czymkolwiek ustalimy, że ona jest.

W swojej wpływowej książce *Vision* (1982) neuronaukowiec i badacz sztucznej inteligencji David Marr zaproponował trzy poziomy analizy wymagane do wyjaśnienia dowolnego zjawiska umysłowego. „Najwyższy", czyli najbardziej abstrakcyjny poziom, obliczeniowy, jest analizą „*problemu* [podkreślenie moje – przyp. D.C.D.] jako zadania przetworzenia informacji", a poziom średni, algorytmiczny, to analiza rzeczywistych procesów, które realizują *to* zadanie przetwarzania informacji. Najniższy poziom, fizyczny, dostarcza analizy maszynerii neuronowej i pokazuje, w jaki sposób wykonuje się algorytmy opisane na poziomie średnim, co tym samym prowadzi do rozwiązania zadania opisanego abstrakcyjnie na poziomie obliczeniowym[86].

Trzy poziomy Marra mogą również zostać użyte do opisu rzeczy o wiele prostszych od umysłów, a różnice między poziomami można zrozumieć, widząc, jak można je zastosować do czegoś tak prostego jak liczydło. Jego zadaniem obliczeniowym jest przeprowadzić obliczenia arytmetyczne: pokazać prawidłowy wynik każdego problemu arytmetycznego przekazanego mu

na wejściu. Na tym poziomie liczydło i kalkulator są więc takie same; są zaprojektowane, aby wykonywać te same „zadania przetwarzania informacji". Algorytmiczny opis liczydła jest tym, czego uczysz się, gdy chcesz się nim posługiwać – przepisem na poruszanie kulek, aby dodać, odjąć, pomnożyć i podzielić. Jego opis fizyczny zależy od tego, z czego jest zrobiony: mogą to być drewniane kulki zawieszone na sznurkach w ramce lub też żetony ułożone wzdłuż linii na podłodze, albo coś stworzonego przez ołówek i dobrą gumkę na kartce papieru w linie.

Marr zalecał modelowanie zjawisk psychologicznych na wszystkich trzech poziomach analizy, a szczególnie podkreślił istotę wyjaśniania poziomu najwyższego, obliczeniowego, zanim ruszy się bez zastanowienia do modelowania na niższych poziomach[87]. Jego badania nad widzeniem w genialny sposób ukazały siłę tej strategii, a inni badacze od tego czasu korzystają z niej przy opisie innych zjawisk. Chciałoby się użyć tych samych trzech poziomów analizy do świadomości jako całości i niektórzy poddali się tej pokusie. Jak jednak widzieliśmy w rozdziale 7, jest to ryzykowne uproszczenie: pytając „Jaka jest funkcja świadomości?", zakładamy, że istnieje jedno „zadanie przetwarzania informacji" (jakkolwiek skomplikowane), do którego wykonania neuronowa maszyneria świadomości jest dobrze zaprojektowana – prawdopodobnie przez ewolucję. Może to prowadzić do przeoczenia istotnych możliwości: że pewne cechy świadomości mają wiele funkcji; że pewne funkcje są słabo wykonywane przez istniejące cechy z powodu historycznych ograniczeń w ich rozwoju; że pewne cechy nie mają żadnej funkcji – a przynajmniej takiej, która służyłaby *naszym* celom. Uważając zatem, aby uniknąć tych przeoczeń, przypomnijmy sobie *możliwości* (niekoniecznie funkcje) mechanizmów opisanych w mojej miniaturze.

Przede wszystkim, jak widzieliśmy w rozdziale 7, pojawiają się niebanalne problemy z samokontrolą ze względu na namnażanie się jednocześnie aktywnych specjalistów, a jednym z fundamentalnych zadań wykonywanych przez maszynę Joyce'owską jest rozstrzyganie sporów, niwelowanie przejść pomiędzy systemami władzy oraz zapobieganie zamachom stanu przez wprowadzanie „odpowiednich" sił. Proste lub przeuczone zadania bez poważnej konkurencji mogą być wykonywane rutynowo bez wzywania dodatkowej pomocy, czyli nieświadomie, gdy jednak zadanie jest trudne czy nieprzyjemne, wymaga „koncentracji", czegoś, co „my" osiągamy dzięki samonapominaniu i różnym innych sztuczkom mnemotechnicznym, ćwiczeniom (Margolis 1989) oraz automanipulacjom (Norman i Shallice 1985). Często odkrywamy, że pomaga nam mówienie na głos, powrót do prymitywnych, choć efektywnych strategii, którymi bezpośrednimi potomkami są nasze prywatne myśli.

Takie strategie autokontroli pozwalają nam rządzić naszymi własnymi procesami percepcyjnymi, otwierając nowe możliwości. Jak zauważył psycholog Jeremy Wolfe (1990), nasze układy wzrokowe są bezpośrednio zaprojektowane do wykrywania pewnego rodzaju obiektów – rodzaju, który „ukazuje się", gdy „po prostu patrzymy"; istnieją inne rodzaje, które możemy zidentyfikować, tylko jeśli ich *szukamy* celowo, w strategii utworzonej przez akt autoreprezentacji. Czerwony punkt w masie punktów zielonych będzie widać jak na dłoni (a raczej będzie się wyróżniał niczym dojrzała jagoda pośród zielonych liści), gdy jednak czyjeś projekty wymagają znalezienia czerwonego punktu w masie punktów innych kolorów, należy zadać *sobie* zadanie szeregowego poszukiwania. Jeśli projektem jest odnalezienie czerwonego, kwadratowego konfetti wśród jego wielokolorowych i wielokształtnych sąsiadów (lub odpowiedź na pytanie „Gdzie jest Wally?" [Handford 1987] w popularnych książkach z obrazkami), zadanie szeregowego przeszukiwania może stać się szczególnie absorbującym, metodycznym projektem, wymagającym wyższego stopnia samokontroli.

Te techniki reprezentowania sobie rzeczy pozwalają być samorządnymi lub kierować samymi sobą w przeciwieństwie do wszystkich innych stworzeń. Możemy z wyprzedzeniem

wypracować strategie dzięki umiejętności hipotetycznego myślenia i tworzenia scenariuszy; możemy wzmocnić własne postanowienia i zająć się nieprzyjemnymi czy długotrwałymi projektami przez nawyki samoprzypominania oraz przemyślenie spodziewanych zysków i strat w przyjętych przez nas strategiach. Co ważniejsze, ta praktyka ćwiczeniowa tworzy pamięć trasy, którą dotarliśmy na miejsce, gdzie się znajdujemy (coś, co psychologowie nazywają pamięcią *epizodyczną*), więc możemy sami sobie wytłumaczyć, gdy znajdziemy się przyciśnięci do muru, jakie błędy popełniliśmy (Perlis 1991). W rozdziale 7 widzieliśmy, jak rozwój tych strategii umożliwił naszym przodkom spoglądanie dalej w przyszłość, a tym, co częściowo dało im tę zwiększoną umiejętność przewidywania, była zwiększona umiejętność przypominania – możliwość spojrzenia dalej w przeszłość na ich własne niedawne aktywności, aby *zobaczyć*, gdzie tkwiły błędy. „Cóż, nie mogę *tego* więcej robić!" – to słowa powtarzane przez każdą istotę uczącą się przez doświadczenie, jednak możemy nauczyć się odsuwać te rzeczy dalej i bardziej odkrywczo niż jakakolwiek inna istota dzięki naszemu nawykowi przechowywania danych – lub też, mówiąc dokładniej, dzięki naszym nawykom autostymulacji, których efektem jest między innymi poprawa pamięci.

Jednak takie obciążanie pamięci jest tylko jednym z cennych efektów tych nawyków. Tak samo ważny jest efekt emitowania (Baars 1988), tworzący swego rodzaju otwarte forum, pozwalając na to, aby *wszystko*, co zostało raz nauczone, miało wpływ na *każdy* bieżący problem. Baars uważa, że ta wzajemna dostępność treści zapewnia *kontekst*, bez którego zdarzenia zachodzące „w świadomości" nie miałyby – *nie mogłyby mieć* – sensu dla podmiotu. Treści tworzące otaczający kontekst nie zawsze są same w sobie świadome – tak naprawdę z reguły nie są w ogóle dostępne, mimo że są aktywowane – jednak połączenia między nimi oraz treści, które mogą się pojawić w relacjach werbalnych, gwarantują coś, co moglibyśmy nazwać ich „świadomie pojmowanym" znaczeniem.

Ray Jackendoff (1987) twierdzi w tym samym duchu, że najwyższe poziomy analizy wykonywanej przez mózg, przez które rozumie te najbardziej abstrakcyjne, *nie* są dostępne w przeżyciu, mimo że umożliwiają przeżywanie, gdyż nadają mu sens. Jego analiza stanowi więc przydatne antidotum na kolejne wcielenie teatru kartezjańskiego jako „szczytu" czy „czubka góry lodowej". (Oto dobry przykład zapożyczony od neuropsychologa Rogera Sperry'ego: „Z pozycji głównodowodzącego na najwyższych poziomach w hierarchii organizacyjnej mózgu, właściwości subiektywne [...] sprawują władzę nad biofizycznymi i chemicznymi czynnościami, odbywającymi się na niższych poziomach [Sperry 1977, s. 117]).

Wielu filozofów, szczególnie ci, na których ma wpływ szkoła fenomenologii Husserla (Dreyfus 1979; Searle 1983), podkreśla znaczenie tego „tła" świadomego przeżycia, jednak zwykle opisują je jako tajemniczą czy krnąbrną cechę, opierającą się wyjaśnieniom mechanicznym, a nie jako klucz, jak sugerują Baars i Jackendoff, do obliczeniowej teorii tych zdarzeń. Owi filozofowie zakładają, że świadomość jest *źródłem* pewnego specjalnego rodzaju „wewnętrznej intencjonalności", ale – jak ujął to filozof Robert van Gulik – oddala nas to od celu.

> Osobisty poziom przeżycia zrozumienia [...] nie jest iluzją. *Ja*, osobisty podmiot przeżycia, rozumiem. Mogę poczynić wszelkie potrzebne połączenia wewnątrz przeżycia, powołując reprezentacje, które natychmiast łączą się ze sobą. Fakt, że moja umiejętność jest rezultatem tego, że składam się ze zorganizowanego systemu części subosobowych składników tworzących mój uporządkowany przepływ myśli, nie zaprzecza istnieniu tej umiejętności. Iluzoryczny lub błędny jest jedynie pogląd, że jestem swego rodzaju osobną jaźnią tworzącą te połączenia dzięki zupełnie niebehawioralnej formie rozumienia. [van Gulik 1988, s. 99]

Wszystko, co wiemy, może mieć wpływ na każdą rzecz, z którą się obecnie zmagasz.

Taki przynajmniej jest ideał. Ta cecha nazywana jest przez Fodora (1983) *izotropią*, umiejętnością, jak powiedziałby Platon, przywołania odpowiednich ptaków, a przynajmniej nakłonienia ich do śpiewania wtedy, gdy jest nam to potrzebne. Wygląda to magicznie, lecz jak wie każdy magik, pozór magii jest zwielokrotniony przez fakt, iż publiczność zwykle ma w zwyczaju wyolbrzymiać zjawisko, próbując je wyjaśnić. Z początku może się wydawać, że jesteśmy idealnie izotropowi, ale tacy nie jesteśmy. Trzeźwe zastanowienie przypomina nam o tych okazjach, gdy wolno rozpoznajemy istotę nowych danych. Pomyśl o klasycznym, komediowym wyolbrzymieniu takiej sytuacji: „spóźniony refleks" (Neisser 1988). Czasem nawet odcinamy gałąź, na której siedzimy, lub odpalamy zapałkę, aby spojrzeć do kanistra pełnego benzyny[88].

Magicy występujący na scenie wiedzą, że zestaw tanich sztuczek często wystarczy do stworzenia „magii", a wie to również Matka Natura, największa kolekcjonerka gadżetów. Badania w dziedzinie sztucznej inteligencji eksplorują możliwe triki, szukając „pakietu [...] heurystycznego, odpowiednio koordynowanego i sprawnie rozmieszczanego" (Fodor 1983, s. 116), mogącego zapewnić izotropię w stopniu charakterystycznym dla nas, ludzi. Modele w rodzaju ACT* i Soar – oraz wiele innych pomysłów rozwijanych w sztucznej inteligencji – są obiecujące, ale nierozstrzygające. Niektórzy filozofowie, szczególnie Dreyfus, Searle, Fodor i Putnam (1988), są pewni, że koncepcja umysłu jako gadżetu jest błędna, i próbują skonstruować argumenty dowodzące niemożliwości takiego zadania (Dennett 1988b, 1991c). Na przykład Fodor twierdzi, że podczas gdy systemy o specjalistycznym przeznaczeniu mogą być wbudowane, to w systemie o przeznaczeniu ogólnym, mogącym wszechstronnie odpowiadać na wszelkie nadchodzące elementy, „może liczyć się *niestabilna, natychmiastowa* łączność" (Fodor 1983, s. 118). Rozpacza nad każdym, kto tworzy taką teorię połączeń, ale nie jest pesymistą: rozpacza *z założenia* (staranny trik). Ma rację, że powinniśmy spodziewać się przybliżenia izotropii dzięki oprogramowaniu, a nie temu, co wbudowane na stałe, lecz jego argument przeciwko hipotezie „wszystkich sztuczek" zakłada, iż jesteśmy lepsi w „uwzględnianiu wszystkiego" niż w rzeczywistości. Jesteśmy w tym nieźli, ale nie fantastyczni. Wyrabiane przez nas nawyki automanipulacji sprawiają, że jesteśmy przebiegłymi badaczami naszych ciężko zdobytych zasobów; nie zawsze udaje nam się namówić odpowiedniego ptaka do śpiewania w odpowiednim momencie, lecz robimy to wystarczająco często, aby ptak był niezłym kompanem.

5. Czy jest to jednak teoria świadomości?

Dotychczas byłem powściągliwy co do świadomości. Uważnie unikałem mówienia, co mówi moja teoria o tym, czym *jest* świadomość. Nie stwierdziłem, że cokolwiek, co przypomina maszynę Joyce'owską, jest świadomością, ani nie powiedziałem, że jakikolwiek konkretny stan takiej wirtualnej maszyny jest stanem świadomym. Powód mojej powściągliwości był taktyczny: chciałem uniknąć sporu o to, czym jest świadomość, dopóki nie dałem sobie szansy pokazania, że przynajmniej wiele z przypuszczanych mocy świadomości może zostać wyjaśnionych przez możliwości maszyny Joyce'owskiej *bez względu na to*, czy maszyna ta realizującemu ją sprzętowi nadaje świadomość.

Czy nie mogłaby istnieć *nie*świadoma istota z wewnętrzną, globalną przestrzenią roboczą, w której demony nadawałyby wiadomości do demonów, tworząc koalicje i całą resztę? Jeśli tak, to wprawiająca w osłupienie ludzka umiejętność sprawnego, wszechstronnego dostosowania stanów umysłowych w odpowiedzi na niemal każdą ewentualność, choćby i nieznaną, nic nie zawdzięcza świadomości, a jedynie obliczeniowej architekturze, która sprawia, że taka

komunikacja jest możliwa. Jeśli świadomość jest czymś więcej niż maszyną Joyce'owską, na razie w ogóle nie przedstawiłem teorii świadomości, nawet jeśli odpowiedziałem na inne skomplikowane pytania.

Dopóki nie stworzyłem całego szkicu teorii, musiałem odsuwać takie wątpliwości na bok, lecz w końcu nadszedł czas, by chwycić byka za rogi i skonfrontować się z samą świadomością, tą całą wspaniałą tajemnicą. Zatem niniejszym deklaruję, że TAK, moja teoria jest teorią świadomości. Ktokolwiek lub cokolwiek ma taką maszynę wirtualną jako swój system sterujący, jest świadomy lub świadome w całym znaczeniu tego słowa i jest świadomy, *ponieważ* ma taką maszynę wirtualną[89].

Jestem teraz gotowy zająć się zarzutami. Możemy zacząć od pytania, na które nie odpowiedziałem dwa akapity wcześniej. Czy coś nieświadomego – na przykład zombi – mogłoby posiadać maszynę Joyce'owską? Pytanie to zakłada zarzut tak niezmiennie popularny w takich momentach, że filozof Peter Bieri (1990) nazwał go *tybetańskim młynkiem modlitewnym.* Pojawia się bez ustanku, bez względu na to, jaka teoria jest przedstawiana:

Rozumiem te wszystkie funkcjonalne szczegóły związane z tym, jak mózg robi to czy tamto, ale potrafię sobie wyobrazić, że *wszystko to* zachodzi w systemie bez pojawienia się prawdziwej świadomości!

Dobrą odpowiedzią, choć rzadko słyszaną, jest: Czy naprawdę potrafisz? Skąd wiesz? Skąd wiesz, że wyobrażasz sobie „to wszystko" wystarczająco szczegółowo i z wystarczającą uwagą na wszystkie konsekwencje? Co sprawia, że uważasz, iż twoje twierdzenie jest przesłanką prowadzącą do jakiejś ciekawej konkluzji? Pomyśl, jak niewzruszeni bylibyśmy, gdyby jakiś współczesny *witalista* powiedział:

Rozumiem te wszystkie rzeczy dotyczące DNA i białek itp., ale potrafię sobie wyobrazić, że odkrywamy jednostkę wyglądającą i zachowującą się zupełnie jak kot, łącznie z krwią w żyłach i DNA w „komórkach", ale nie żyjącą. (Czy naprawdę mogę? Pewnie: słyszę miauczenie, a potem Bóg szepcze mi do ucha: „To nie jest żywe! To tylko taki mechaniczny wihajster DNA!". W mojej wyobraźni wierzę Mu).

Ufam, że nikt nie uważa, że jest to dobry argument za witalizmem. Ten wysiłek wyobraźni się nie liczy. Dlaczego? Gdyż jest zbyt wątły w zetknięciu ze sprawozdaniem z życia przedstawianym przez współczesną biologię. Jedyne, co pokazuje ten „argument", to że można zignorować „to wszystko" i uczepić się tego przekonania, jeśli ma się wystarczającą determinację. Czy tybetański młynek modlitewny jest lepszy jako argument przeciwko przedstawionej przeze mnie teorii?

Możemy teraz, dzięki całemu rozciąganiu wyobraźni w poprzednich rozdziałach, przesunąć ciężar dowodu. Tybetański młynek modlitewny (a jak zobaczymy, istnieje kilka jego odmian) jest potomkiem słynnego argumentu Kartezjusza (zob. rozdział 2), w którym twierdzi on, iż jest w stanie pojąć *jasno i wyraźnie*, że jego umysł różni się od mózgu. Siła takiego argumentu istotnie zależy od tego, jak wysokie mamy standardy pojmowania. Niektórzy mogą twierdzić, że są w stanie jasno i wyraźnie pojąć największą liczbę pierwszą lub trójkąt niebędący figurą sztywną. Są w błędzie – a przynajmniej to, co robią, mówiąc, że pojmują te rzeczy, nie powinno być brane za oznakę tego, co możliwe. Obecnie możemy sobie wyobrazić „to wszystko" dosyć szczegółowo. Czy *naprawdę* potrafisz sobie wyobrazić zombi? Jedyny sens, w jakim jest oczywiste, że potrafisz, to nie sens podważający moją teorię, lecz silniejszy, niejasny sens wymagający dowodu.

Filozofowie z zasady tego nie wymagają. Najbardziej wpływowe eksperymenty myślowe we współczesnej filozofii umysłu zakładają zaproszenie publiczności do wyobrażenia sobie jakiegoś specjalnie wymyślonego czy określonego stanu rzeczy, a następnie – bez

odpowiedniego sprawdzenia, czy to zadanie świadomości zostało rzeczywiście wypełnione – zaproszenie publiczności do „zwrócenia uwagi” na różne konsekwencje tej fantazji. Te „dźwignie wyobraźni”, jak je nazywam, są piekielnie sprytnymi mechanizmami. Zasługują na swoją sławę, nawet jeśli tylko z powodu własnego uwodzicielskiego charakteru.

Zajmiemy się nimi w części III, po drodze rozwijając naszą teorię świadomości. Z naszej nowej perspektywy będziemy w stanie zobaczyć zręczność, z jaką mylnie kierują publiczność – oraz iluzjonistów – a w międzyczasie wyostrzymy naszą własną wyobraźnię. Pośród słynnych argumentów odnajdziemy nie tylko przypuszczalną możliwość istnienia zombi, ale także odwrócone spektrum, czego badaczka kolorów Maria nie wie o kolorach, chiński pokój oraz jak to jest być nietoperzem.

Część trzecia

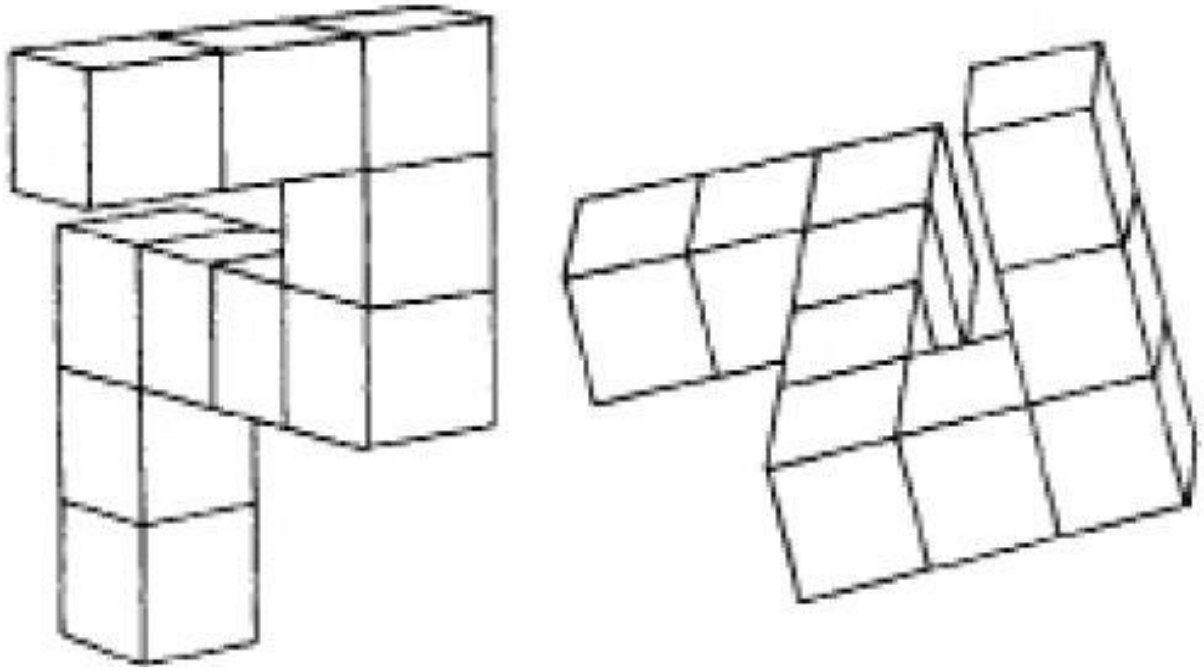
Filozoficzne problemy świadomości

Rozdział 10

Pokaż i powiedz

1. Obracanie obrazów oczyma wyobraźni

Pierwsze wyzwanie, zanim zajmiemy się filozoficznymi eksperymentami myślowymi, pochodzi z pewnych realnych eksperymentów mogących się wydawać rehabilitacją teatru kartezjańskiego. Niektóre najbardziej ekscytujące i genialne badania w kognitywistyce w ciągu ostatnich dwudziestu lat dotyczyły ludzkiej umiejętności manipulowania obrazami umysłowymi, a zapoczątkował je psycholog Roger Shepard (Shepard i Metzler 1971) w swoim klasycznym badaniu *prędkości umysłowej rotacji* brył takich jak te na rycinie 10.1.



Ryc. 10.1

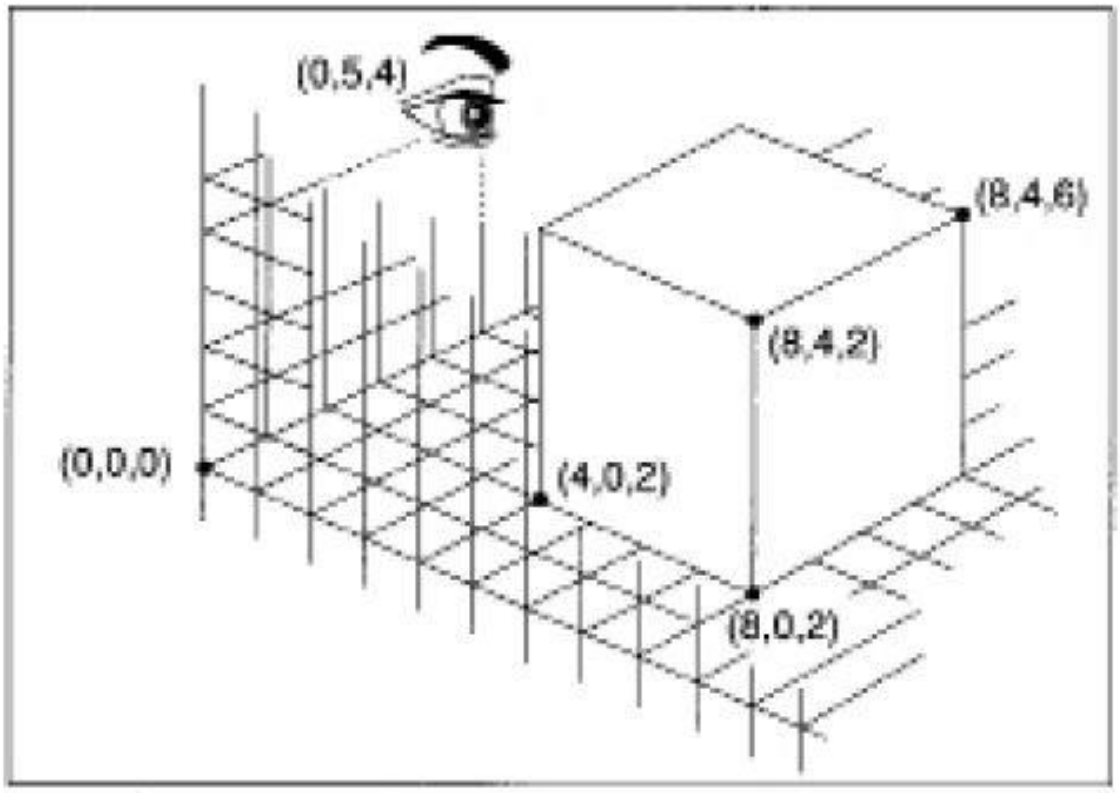
W oryginalnym eksperymencie pokazano badanym pary rysunków liniowych i zapytano ich, czy są one tym samym kształtem przedstawionym z różnych perspektyw, czy nie. W powyższym przypadku, jak szybko możesz stwierdzić, odpowiedź brzmi „tak". Skąd wiesz? Typowa odpowiedź to: „Obróciłem jeden z obrazków oczyma wyobraźni, po czym nałożyłem go na drugi". Shepard przygotowywał pary brył obróconych o różne odległości kątowe – niektóre z nich były pokazane w perspektywach jedynie o kilka stopni różnych od siebie, a inne trzeba było bardzo obrócić, aby ustawić je pod tym samym kątem – i mierzył średni czas, którego potrzebowali badani na odpowiedź dotyczącą każdej z par. Zakładając, że coś w rodzaju obracania obrazu w mózgu rzeczywiście występuje, obrócenie obrazu umysłowego o 90 stopni powinno zająć dwa razy więcej czasu niż obrócenie go o 45 stopni (jeśli nie bierzemy pod uwagę przyspieszenia i zwolnienia, zachowując stałą prędkość obrotu)[90]. Dane Sheparda w sposób

niezwykły potwierdziły tę hipotezę w różnorodnych warunkach. Setki eksperymentów przeprowadzonych później, przez Sheparda i innych, badały zachowanie maszynerii obracającej obrazy mózgowe bardzo dokładnie i – przedstawiając nadal sporny konsensus jak najostrożniej – rzeczywiście wydaje się, że w mózgu istnieje coś, co psycholog Stephen Kosslyn (1980) nazywa „buforem wizualnym", który wykonuje transformacje poprzez procesy będące silnie „wyobrażeniowe" – czy, używając pojęcia Kosslyna, *quasi-obrazkowe*.

Cóż to oznacza? Czy psychologowie kognitywni odkryli, że teatr kartezjański tak naprawdę istnieje? Według Kosslyna te eksperymenty pokazują, że obrazy są składane w celu wyświetlania wewnętrznego obrazu, bardzo podobnie do obrazów na kineskopie (takim jak ekran telewizora czy komputera) tworzonych z dokumentów w pamięci komputera. Gdy znajdą się już na wewnętrznym ekranie, mogą być obracane, przeszukiwane i w inny sposób zmieniane przez osoby badane, które otrzymały konkretne zadanie do wykonania. Kosslyn podkreśla jednak, że jego model kineskopu jest metaforą. Powinno nam to przypomnieć metaforyczny talent Shakeya do „operowania obrazami". Shakey z pewnością nie miał teatru kartezjańskiego w swoim komputerowym mózgu. Chcąc jaśniej zrozumieć, co rzeczywiście musi się odbywać w ludzkim mózgu, możemy zacząć od modelu *nie*metaforycznego, zbyt mocnego, aby okazał się prawdą, a następnie po kolei „odejmować" od niego niechciane właściwości. Innymi słowy, zajmiemy się metaforą kineskopu Kosslyna i stopniowo wprowadzimy w nim ograniczenia.

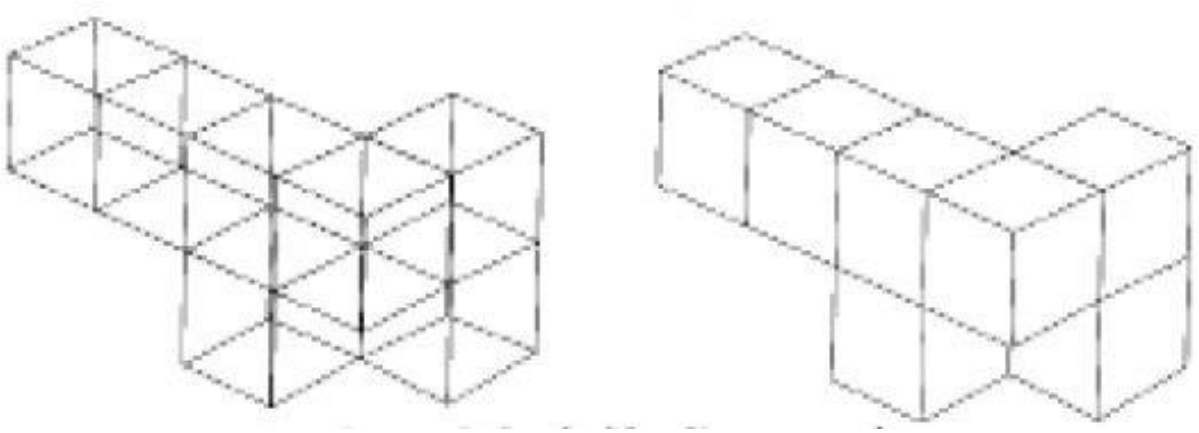
Najpierw wyobraźmy sobie system, który rzeczywiście operuje prawdziwymi obrazami, taki jak graficzne systemy komputera występujące teraz w setkach układów: komputerowe animacje w telewizji i filmach, systemy przedstawiające trójwymiarowe przedmioty z różnej perspektywy dla użytku architektów i dekoratorów wnętrz, gry wideo i wiele innych. Inżynierowie nazywają ich wersje systemami CAD, czyli komputerowym wspomaganiem projektowania Systemy CAD rewolucjonizują inżynierię nie tylko dlatego, że ułatwiają szkicowanie, tak jak edytory tekstu upraszczają pisanie, ale ponieważ inżynierowie mogą szybko rozwiązywać problemy i odpowiadać za ich pomocą na pytania, które w innym razie byłyby dosyć skomplikowane. Inżynier, napotykając problem Sheparda z ryciny 10.1, mógłby odpowiedzieć na zadane pytanie za pomocą systemu CAD, umieszczając oba obrazy na ekranie CAD i dosłownie obracając jeden z nich, a następnie nakładając go na drugi. Kilka szczegółów tego procesu jest istotnych.

Każdy z przedstawionych na obrazku obiektów zostałby wprowadzony do pamięci komputera jako trójwymiarowy obiekt *wirtualny*, zredukowany do opisu powierzchni i krawędzi zdefiniowanych za pomocą współrzędnych *xyz*, a każdy z elementów zajmowałby punkt w przestrzeni wirtualnej jako „uporządkowana trójka" liczb przechowywanych w pamięci komputera. Punkt widzenia zakładanego obserwatora również zostałby wprowadzony jako punkt w tej samej przestrzeni wirtualnej, zdefiniowany przez swoją własną uporządkowaną trójkę współrzędnych. Poniżej widać diagram przedstawiający sześcian i punkt widzenia, lecz warto pamiętać, że komputer musi przechowywać tylko takie trójki dla każdego kluczowego punktu, złożone w większe grupy (np. każda ze ścian sześcianu), łącznie z zakodowanymi informacjami dotyczącymi różnych właściwości każdej ze ścian (jej kolor, to, czy jest przejrzysta, czy nie, jej faktura itp.). Obliczenia związane z obracaniem jednego z obiektów i przesuwaniem go w wirtualnej przestrzeni mogą być szybko wykonane przez proste dostosowanie współrzędnych *x*, *y* i *z* tego obiektu – dzięki banalnej arytmetyce. Następnie jest kwestią prostej geometrii wyliczenie linii determinujących, które powierzchnie obiektu będą widziane z wirtualnego punktu widzenia i dokładnie jak będą się prezentować. Takie wyliczenia są łatwe, ale pracochłonne lub „drogie obliczeniowo", szczególnie jeśli wyliczone muszą być również gładkie wygięcia, zacieniowanie, odbite światło czy faktura.



Ryc. 10.2

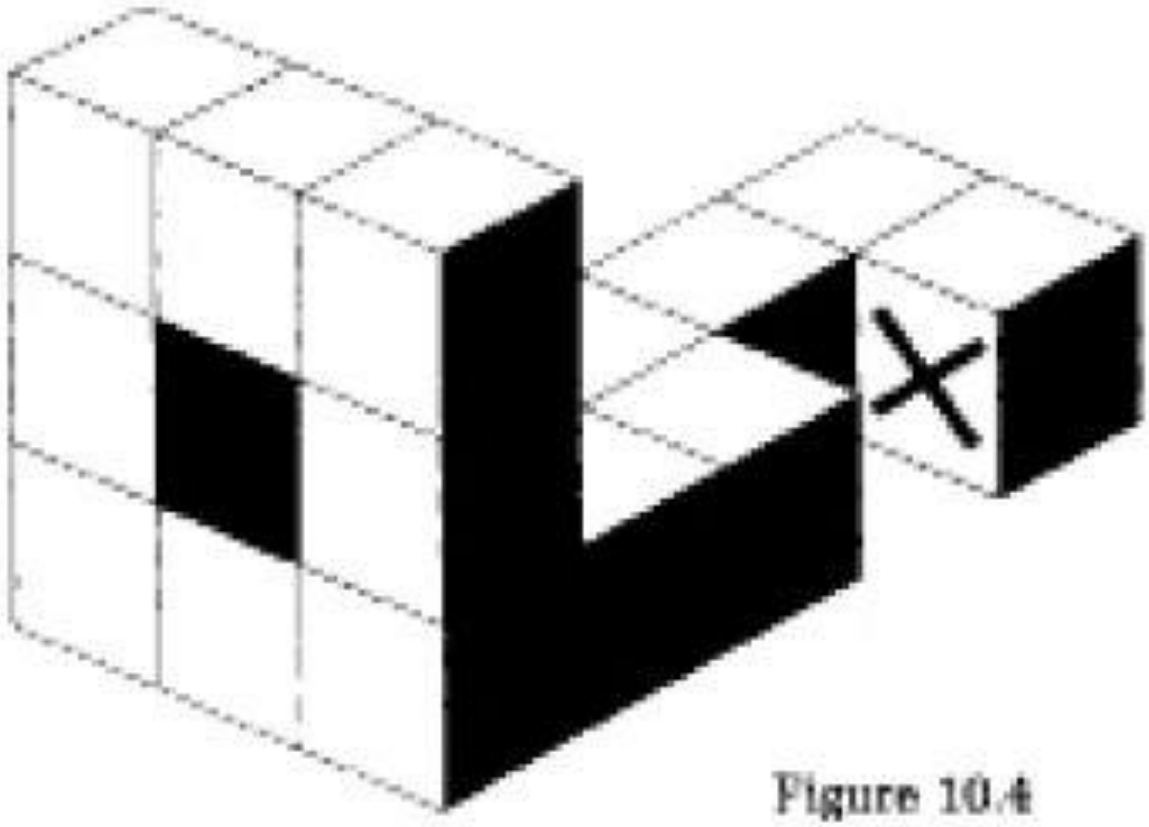
W systemach zaawansowanych różne klatki mogą być obliczone wystarczająco szybko, aby stworzyć pozorny ruch na ekranie, ale tylko jeśli reprezentacje są w miarę schematyczne. „Ukryte usuwanie linii", proces obliczeniowy przedstawiający ostateczny obraz jako nieprzejrzysty w odpowiednich miejscach i zapobiegający temu, aby sześcian Sheparda wyglądał jak przezroczysty sześcian Neckera, jest sam w sobie względnie czasochłonnym procesem, który mniej więcej wyznacza ograniczenia dla tego, co może być przedstawione w „czasie rzeczywistym". Aby otrzymać zachwycająco szczegółowy obraz transformacji, który oglądamy codziennie w telewizji, za pomocą grafiki komputerowej, proces generacji obrazu musi być o wiele wolniejszy, nawet na superkomputerach, a pojedyncze klatki muszą być przechowywane w celu późniejszego zaprezentowania w szybszym tempie, spełniającym wymagania wykrywania ruchu przez ludzki system wzrokowy[91].



Ryc. 10.3
Przed i po usunięciu ukrytych linii

Takie manipulatory trójwymiarowych obiektów wirtualnych są wspaniałymi, nowatorskimi narzędziami lub zabawkami i rzeczywiście są czymś nowym pod słońcem, a nie elektroniczną kopią czegoś, co już mamy w naszych głowach. Jest raczej pewne, że żaden proces analogiczny do tych bilionów geometrycznych i arytmetycznych kalkulacji nie odbywa się w naszym mózgu, gdy zajmujemy się umysłowym obrazowaniem, i nic innego *nie mogłoby* pokazać szczegółowych sekwencji animacyjnych, które wytwarzają – z powodów, o których mówiliśmy w rozdziale 1.

Możemy się pocieszyć tym, że to ograniczenie naszych mózgów jest rzeczywiste, rozważając nieco inny problem w rodzaju Sheparda, który byłby w miarę łatwy do rozwiązania za pomocą takiego systemu CAD: Czy czerwone *X* na jednym z boków tego obiektu jest widoczne dla kogoś patrzącego przez dziurę w jego przedniej ścianie?



Ryc. 10.4

Nasz obiekt Sheparda z *X* jest prosty i schematyczny, a skoro pytanie, na które chcemy odpowiedzieć, jest niezależne od faktury, oświetlenia i innych tego rodzaju subtelności, dla inżyniera wytworzenie animowanego obrotu tego obiektu na kineskopie byłoby stosunkowo łatwe. Mógłby potem obrócić obraz w każdą stronę, przesuwając punkt widzenia tam i z powrotem – a następnie po prostu *szukać* mignięcia koloru czerwonego przez dziurę. Jeśli zobaczy czerwony, odpowiedź brzmi „tak", w innym przypadku – „nie".

Czy potrafisz przeprowadzić ten sam eksperymenty oczyma wyobraźni? Czy możesz po prostu obrócić pokazany obiekt i spojrzeć przez dziurę? Jeśli tak, to znaczy, że potrafisz zrobić coś, czego ja nie potrafię, a wszyscy ludzie, których o to prosiłem, również nie byli w stanie tego zrobić. Nawet ci, którzy mają odpowiedź na zadane pytanie, są raczej pewni, że nie odnaleźli jej *jedynie* poprzez obrócenie obiektu i spojrzenie przez dziurę. (Często mówią, że najpierw próbowali obracania i spoglądania i stwierdzili, że to nie działa; mogli „obrócić obiekt", ale „rozpadł się on", gdy próbowali spojrzeć przez dziurę. Następnie mówią o rysowaniu linii wzroku przechodzącej przez dziurę na nieobróconym obrazie, aby zobaczyć, czy są w stanie stwierdzić, w którym miejscu te linie spotkałyby się z tylną powierzchnią). Nasz obiekt Sheparda nie jest bardziej skomplikowany niż obiekty, które najwyraźniej zostały pomyślnie obrócone w wielu eksperymentach, więc pojawia się zagadka: Jakiego rodzaju proces może tak szybko dokonać pewnych transformacji (następnie zaś uzyskać informacje z ich wyniku), a jednocześnie tak nie radzić sobie z innymi operacjami, które nie wydają się trudniejsze? (Jeśli te operacje nie *wydają* się nam trudniejsze, najwyraźniej patrzymy na nie ze złego punktu widzenia, ponieważ nasza porażka pokazuje, że są one bardziej wymagające).

Eksperyment przeprowadzony przez psychologów Daniela Reisberga i Deborah Chambers (1991) prowadzi do tego samego pytania. Badanym, którzy twierdzili, że mają dobrą wyobraźnię, pokazano „bezsensowne" kształty. Następnie poproszono ich o obrócenie umysłowe kształtów o 90 lub 180 stopni w wyobraźni i powiedzenie, co „widzieli". Byli zaskoczeni, że nie byli w stanie rozpoznać oczyma swojej wyobraźni tego, co bardzo szybko rozpoznajemy, gdy obrócimy książkę zgodnie z ruchem wskazówek zegara o 90 stopni i spojrzymy na kształty.

Ryc. 10.5

Inżynierowie używają systemów CAD do odpowiedzi na pytania, które nie są zwykle tak proste, jak „Czy czerwone X jest widoczne przez dziurę?". Zazwyczaj są one związane z bardziej skomplikowanymi właściwościami przestrzennymi projektowanych przedmiotów, takimi jak „Czy ten robot z ręką o trzech stawach będzie w stanie dotrzeć do pokrętła na swoich plecach bez zderzenia z zasilaczem?", a nawet bardziej estetyczne właściwości takich obiektów, na przykład: „Jak będą wyglądały schody w lobby tego hotelu dla kogoś idącego ulicą i patrzącego przez szklane okna?". Gdy próbujemy zwizualizować takie sceny bez żadnej pomocy, otrzymujemy jedynie ogólnikowe i zawodne rezultaty, więc system CAD może być postrzegany jako rodzaj protezy wyobraźni (Dennett 1982d, 1990b). Znacznie poszerza on moce wyobraźni istoty ludzkiej, ale jest zależny od normalnego widzenia użytkownika – aby ten mógł spojrzeć na kineskop.

Teraz spróbujmy sobie wyobrazić bardziej ambitne urządzenie-protezę: system CAD dla niewidomych inżynierów! Aby zbytnio nie komplikować, załóżmy, że pytania, na które odpowiedzi szukają owi niewidomi inżynierowie, są stosunkowo proste, geometryczne – nie są natomiast pytaniami o estetykę architektury. Informacje wyjściowe będą oczywiście musiały mieć postać niewizualną. Formą najłatwiejszą w użytku byłyby zwykłe odpowiedzi językowe (w brajlu lub przekazane przez syntezator dźwięku) na pytania językowe. Załóżmy więc, że w sytuacji, gdy niewidomy inżynier będzie potrzebował odpowiedzi na pytania z rodzaju tego, o jakim mówiliśmy przed chwilą, przekaże on po prostu zdanie do systemu CAD (oczywiście w sposób dla niego „zrozumiały") i będzie czekał, aż system CAD udzieli odpowiedzi.

Nasz system CAD Dla Niewidomych 1.0 jest mało elegancki, ale prosty. Składa się ze zwykłego systemu CAD, łącznie z CRT, przed którym znajduje się wizualny system komputera – *Vorsetzer* – złożony z kamery skierowanej na kineskop oraz palców robota mających poruszać gałkami systemu CAD[92]. W przeciwieństwie do Shakeya, u którego kineskop istniał jedynie dla użytku obserwatorów, ten system rzeczywiście „patrzy na" obraz, prawdziwy obraz złożony ze świecących kropek, które wypromieniowują prawdziwe światło o różnych częstotliwościach na

wrażliwe na światło przetworniki znajdujące się z tyłu kamery. Gdy natknie się na nasz problem czerwonego X Sheparda, CAD Dla Niewidomych 1.0 stworzy obraz z prawdziwym, czerwonym X, widocznym dla wszystkich, również dla kamery Vorsetzera.

Przejdźmy do rzeczy i załóżmy, że Vorsetzer rozwiązał sam w sobie wystarczająco wiele problemów maszynowym postrzeganiem obrazu, aby móc wydobyć poszukiwane informacje z reprezentacji świecących na ekranie kineskopu. (Nie, nie zamierzam twierdzić, że Vorsetzer jest świadomy – chcę jedynie założyć, że jest wystarczająco dobry w tym, co robi, aby móc odpowiedzieć na pytania stawiane mu przez niewidomego inżyniera). CAD Dla Niewidomych 1.0 tworzy prawdziwe obrazy, manipuluje nimi i używa ich do odpowiadania niewidomemu inżynierowi na pytania, na które widzący inżynier mógłby odpowiedzieć, korzystając ze zwykłego systemu CAD. Jeśli system 1.0 jest tak dobry, to system 2.0 będzie banalny do zaprojektowania: po prostu wyrzucamy kineskop i skierowaną nań kamerę, a na ich miejsce wprowadzamy prosty kabel! Przez ten kabel system CAD przesyła Vorsetzerowi *mapę bitową*, ciąg zer i jedynek definiujący obraz na kineskopie. W Vorsetzerze systemu 1.0 ta mapa bitowa była mozolnie rekonstruowana z danych wyjściowych przetworników optycznych w kamerze.

Niewiele jest oszczędności w *obliczeniach* w systemie 2.0 – następuje jedynie eliminacja niepotrzebnego sprzętu. Wszystkie rozbudowane wyliczenia linii wzroku, usuwanie ukrytych linii i tworzenie faktury, cieni oraz odbijającego się światła, które wymagały tyle obliczeń w systemie 1.0, nadal są częścią procesu. Wyobraźmy sobie, że Vorsetzer w systemie 2.0 ma dokonać głębokiej oceny przez porównanie kąta nachylenia faktury czy interpretację cienia. Będzie musiał *zanalizować* schematy bitów na odpowiednich pozycjach mapy bitowej, aby dotrzeć do rozróżnień faktur i cieni.

Oznacza to, że system 2.0 jest nadal absurdalnie niewydajną maszyną, ponieważ informacja o tym, że jakiś konkretny fragment mapy bitowej powinien reprezentować cień, jest już „znana" systemowi CAD (jeśli jest to część zakodowanego opisu obiektu, z którego system CAD generuje swoje obrazy), a jeśli ten fakt Vorsetzer musi dopiero ustalić, aby móc poczynić głęboką ocenę, dlaczego system CAD po prostu tego Vorsetzerowi nie *powie*? Po co zajmować się *prezentowaniem* cienia w celu analizowania schematów w Vorsetzerze, jeśli zadanie prezentowania schematu i jego analizowania wzajemnie się eliminują?

Zatem nasz CAD Dla Niewidomych 3.0 będzie wolny od ogromnych zadań obliczeniowych związanych z prezentacją obrazu, biorąc większość tego, co „wie" o reprezentowanych obiektach, i przekazując te informacje bezpośrednio do podsystemów Vorsetzera, a przy okazji korzystając z formatów prostych kodów właściwości i przypinając „etykiety" różnym „obszarom" na sieci mapy bitowej, co następnie zostaje zamienione z czystego obrazu na coś w rodzaju diagramu. *Niektóre* właściwości przestrzenne są reprezentowane bezpośrednio – są *pokazywane* – w (wirtualnej) przestrzeni mapy bitowej, ale inne są jedynie *opowiedziane* przez etykiety[93].

Powinno nam to przypomnieć o moim twierdzeniu z rozdziału 5, że mózg musi poczynić dane rozróżnienie tylko raz; zidentyfikowana cecha nie musi być zaprezentowana ponownie dla centralnej istoty doceniającej w teatrze kartezjańskim.

Teraz jednak widzimy inny aspekt inżynierii: „wzajemne eliminowanie się" działa tylko wówczas, gdy systemy zmuszone się komunikować „mówią tym samym językiem". Co jeśli format, w którym system CAD już „zna" właściwe informacje – na przykład dane o tym, że coś jest cieniem – nie jest formatem, w którym tej informacji może „użyć" Vorsetzer?[94] Wówczas komunikacja może wymagać „cofnięcia się, aby skoczyć do przodu". Może być konieczne, aby systemy zaangażowały się w informacyjnie rozrzutne – można by rzec: rozwlekłe – interakcje, by w ogóle miały kontakt. Przypomina to szkicowanie mapy, żeby wytłumaczyć drogę

obcokrajowcowi, podczas gdy musi on wiedzieć – gdybyśmy umieli to powiedzieć w jego języku – tylko „skręć w lewo na najbliższych światłach". Zawracanie sobie głowy tworzeniem czegoś na kształt obrazu jest często potrzebne dla celów praktycznych, nawet jeśli nie jest potrzebne „co do zasady".

Skoro systemy w naszych mózgach są wytworami kilku nakładających się na siebie historii oportunistycznego majsterkowania, długiej historii doboru naturalnego i krótkiej historii indywidualnego przekonstruowania poprzez automanipulację, powinniśmy się spodziewać, że odnajdziemy takiego rodzaju niewydajność. Poza tym istnieją inne powody przedstawiania informacji w formatach obrazkowych (poza czystą przyjemnością robienia tego), które – jeśli przypadkiem się na nie natkniemy – wkrótce zrobią na nas wrażenie, sprawiając, że tworzenie obrazów będzie w każdym razie warte zachodu. Jak już zauważyliśmy w spekulacjach w rozdziale 7 dotyczących „tworzenia diagramów dla samego siebie", tego rodzaju transformacje formatu są często niezwykle efektywnymi sposobami wydobywania informacji, która w innym razie jest zupełnie nie do uzyskania z danych. Diagramy rzeczywiście sprowadzają się do re-*prezentacji* informacji – nie dla wewnętrznego oka, ale dla wewnętrznego mechanizmu rozpoznawania wzorców, który również przyjmuje dane wejściowe ze zwykłego („zewnętrznego"?) oka. Dlatego techniki grafiki (komputerowej) są tak cenne na przykład dla nauki. Pozwalają zaprezentować ogromne tablice danych w formie dostępnej dla znakomitych umiejętności rozpoznawania wzorców obecnych w ludzkim oku. Tworzymy wykresy, mapy i wszelkiego rodzaju barwnie kodowane diagramy po to, aby poszukiwane regularności i cechy charakterystyczne po prostu nam się „ukazały" dzięki naszym systemom wzrokowym. Diagramy nie tylko pomagają nam dojrzeć wzory, które w innym przypadku mogłyby być niedostrzegalne; mogą nam ułatwić *śledzić* to, co istotne, oraz *przypominać* nam, aby zadać odpowiednie pytania w odpowiednich momentach. Lars-Erik Janlert (1985), szwedzki badacz sztucznej inteligencji, uważa, że takie generowanie i badanie obrazów w komputerze może też być przydatne w rozwiązywaniu inaczej nierozwiązywalnych problemów dotyczących tego, co moglibyśmy nazwać „zarządzaniem-wnioskowaniem" w systemach, które są „zasadniczo" systemami czysto dedukcyjnymi. (Inny punkt widzenia na ten sam proces przedstawiają Larkin i Simon 1987).

Ta taktyka jest z pewnością dobrze znana wielu bystrym myślicielom i została wspaniale opisana przez jednego z najbystrzejszych, fizyka Richarda Feynmana w *Pan raczy żartować, panie Feynman!*. W rozdziale celnie zatytułowanym *Inny przybornik* opowiada nam, jak zaskoczył swoich kolegów doktorantów w Princeton, „wyczuwając intuicyjnie" prawdę i fałsz tajemnicy teorii topologicznych, których zupełnie nie był w stanie otrzymać formalnie, a nawet w pełni zrozumieć:

Miałem pewną metodę, którą do dziś stosuję, kiedy ktoś stara się coś mi wytłumaczyć: krok po kroku wyobrażam sobie przykłady. Matematycy podawali przykład jakiegoś genialnego twierdzenia, którym się strasznie ekscytowali. Gdy wymieniali założenia, ja budowałem sobie konstrukcję, która je wszystkie spełniała. Gdy była mowa o zbiorze, podstawiałem sobie w głowie piłkę, gdy o zbiorach rozłącznych – dwie piłki. Potem, w miarę przybywania warunków, piłki przybierały w mojej głowie różne kolory, porastały włosami *et cetera*. Potem matematycy recytowali jakieś durne twierdzenie, które nie było prawdziwe dla mojej włochatej zielonej piłki, więc mówiłem: „Fałszywe!".

Jeżeli twierdzenie było prawdziwe, strasznie się podniecali, a ja przez chwilę pozwalałem im się nacieszyć, po czym podawałem im mój kontrprzykład.

– Ach, zapomnieliśmy ci powiedzieć, że to homomorficzny Hausdorff drugiej klasy.

– W takim razie to trywialne! – odpowiadałem. – Trywialne! – Wtedy już kojarzyłem, o co w twierdzeniu chodziło, chociaż nie miałem pojęcia, co to jest homomorficzny Hausdorff

drugiej klasy. [Feynman 1985/2007, s. 89–90]

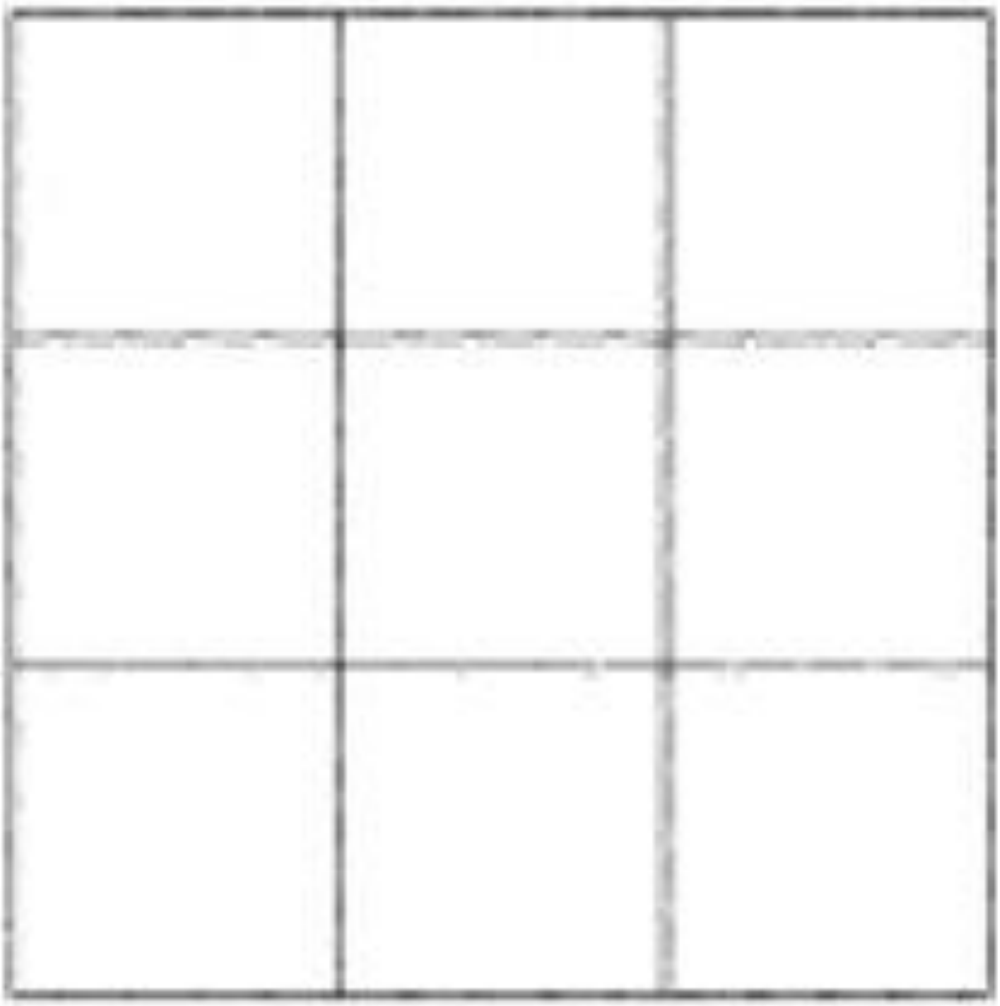
Takie taktyki do pewnego stopnia „przychodzą naturalnie", ale muszą zostać wyuczone lub odkryte, a niektórzy ludzie dają sobie radę o wiele lepiej niż inni. Ci, u których takie umiejętności są wysoce rozwinięte, mają inne maszyny wirtualne w swoich mózgach, o znacząco większym potencjale w porównaniu z tymi, którzy są rzadkimi i marnymi „wizualizatorami". Te różnice szybko ujawniają się w ich indywidualnych, heterofenomenologicznych światach.

Istnieje więc istotny powód, aby uważać, jak twierdzą Kosslyn i inni, że ludzie wykorzystują swoje systemy wizualne nie tylko poprzez prezentowanie sobie prawdziwych, *zewnętrznych* obrazów (jak na kineskopie w systemie CAD), ale również z indywidualnie zaprojektowanymi wewnętrznymi obrazami wirtualnymi czy reprezentacjami danych w formie diagramów, które są odpowiednim surowcem dla jakiegoś późniejszego etapu bądź etapów maszynerii przetwarzania wizualnego.

Na jakie inżynieryjne rozwiązania problemów wewnętrznej komunikacji i manipulacji informacją wpadł ludzki mózg oraz jakie są ich mocne i słabe strony?[95] Są to empiryczne pytania, na które odpowiadają badania nad wyobrażeniami w psychologii poznawczej, i powinniśmy zachować ostrożność, jeśli chodzi o przedstawianie odpowiedzi *a priori*[96]. Przypuszczam, że moglibyśmy znaleźć systemy przekształceń wyobrażeń systemu 1.0 w naszym mózgu, łącznie ze świecącymi kropkami fosforowymi oraz wewnętrznymi oczami wrażliwymi na światło. (Według mnie nie jest *niemożliwe*, aby stworzenia na jakiejś planecie miały takie urządzenie). A eksperymenty takie jak Reisberga i Chambersa pokazują, że skróty, jakie odnalazły nasze mózgi, właściwie uniemożliwiają odkrycie systemu 2.0 z formatem mapy bitowej, która nigdy nie wykorzystuje skrótów. (Gdybyśmy mieli taki system, rozwiązanie problemu czerwonego *X* w naszych głowach byłoby łatwe, tak jak obrócenie Teksasu).

Fenomenologia dostarcza tropów biegnących w obu kierunkach: „mglistość" obrazów mentalnych, która jest „intuicyjnie oczywista" w fenomenologii większości osób badanych, wskazuje na skróty używane przez mózg przypadki, kiedy mózg mówi bez pokazywania. Jest to prawda zarówno w przypadku *percepcji* wizualnej, jak i *wizualizacji*. Zauważyliśmy już w rozdziale 3, jak trudno narysować różę znajdującą się wprost przed naszymi oczami, a nawet skopiować rysunek, a powód jest taki, że czysto przestrzenne właściwości, które jednostka musi zidentyfikować lub rozróżnić, aby rysować dobrze, w toku przetwarzania percepcyjnego zostawały zwykle porzucane, nie są podsumowywane w raportach ani pozostawiane do późniejszego zinterpretowania. Z drugiej strony przydatność obrazów umysłowych w „dostrzeganiu wzorców" lub „przypominaniu sobie" o szczegółach, o których w innym przypadku moglibyśmy zapomnieć, wskazuje na wykorzystanie maszynerii rozpoznawania wzorców wizualnych, które mogło następować, jeśli jedna część mózgu zadała sobie trud przygotowania wersji w specjalnym formacie informacji użytku tych systemów wizualnych. Jak widzieliśmy w rozdziale 1, wymagania związane z przetwarzaniem informacji w takich reprezentacjach są ogromne i nie powinno nas dziwić, że tak kiepsko radzimy sobie z utrzymywaniem w stabilności nawet wysoce schematycznych diagramów w naszych głowach.

Oto krótki test, który przypomni ci o tym, jak bardzo ograniczone są nasze możliwości: *Oczyma wyobraźni* uzupełnij następującą krzyżówkę, wpisując w nią następujące słowa *pionowo* w kolumnach, zaczynając od kolumny lewej: *TOK IWO RYT*.



Ryc. 10.6

Czy potrafisz sprawnie *odczytać* słowa poziome? W rzeczywistym diagramie na papierze te słowa by „wyskoczyły” – nie sposób ich *nie* zobaczyć. Taki w końcu jest cel kreślenia diagramów: przedstawić dane w formacie, który sprawia, że nowa analiza danych staje się łatwa lub nieunikniona. Siatka liter alfabetu trzy na trzy nie jest bardzo skomplikowaną strukturą danych, ale najwyraźniej nie jest czymś, co nasze mózgi potrafią utrzymać w miejscu wystarczająco stabilnie, aby ich systemy wizualne mogły sprawić, że coś „wyskoczy”. (Jeśli chcesz spróbować jeszcze raz, oto dwie kolejne grupy słów do kolumn: *LUJ ELA WEK* i *ZUP ELA ZEW*).

Jest jednak mnóstwo możliwości wariacji indywidualnych w taktykach wykorzystywanych przez różne osoby wizualizujące, a niektóre mogą być w stanie odkryć – lub rozwinąć – strategie wyobrażeniowe pozwalające im „czytać” z tych diagramów. Mając wyjątkowe talenty obliczeniowe, można nauczyć się mnożyć przez siebie liczby dziesięciocyfrowe, więc nie byłoby zaskoczeniem, gdyby niektórzy mogli rozwinąć niezwykłe talenty „czytania krzyżówek” oczyma wyobraźni. Te nieformalne demonstracje dają nam

podpowiedzi, lecz eksperymenty mogą o wiele wyraziściej zdefiniować rodzaje mechanizmów i procesów, z których ludzie korzystają w tych aktach automanipulacji. Dotychczasowe świadectwa wspierają pogląd, że używamy strategii mieszanej, korzystając z wizualnej analizy sieci, jak również włączając etykiety skrótowe, *mówiące*, a nie *pokazujące*.

Zwróćmy jednak uwagę, że nawet w systemie CAD Dla Niewidomych 2.0, który jest skrajnie obrazkowy, po dodaniu mapy bitowej przedstawiającej kolor, zacieniowanie oraz fakturę piksel po pikselu nadal istnieje sens – sens istotny metafizycznie, jak zobaczymy w następnych dwóch rozdziałach – w którym wszystko jest „powiedziane", a nie „pokazane". Przypomnijmy sobie czerwone *X* na naszej figurze Sheparda (Ryc. 10.4). W systemie 1.0 jest przedstawione w prawdziwej czerwieni – kineskop emituje światło, które musi zostać przeniesione do kamery w sposób analogiczny do tego, jak czopki w naszych oczach reagują na różną częstotliwość fal. Gdy Vorsetzer obraca obraz tam i z powrotem, polując na mignięcie czerwieni w dziurze, czeka na okrzyk swoich demonów wykrywających kolor czerwony. W systemie 2.0 to oprzyrządowanie zostaje wyrzucone, a mapa bitowa reprezentuje kolor każdego piksela numerycznie. Być może odcień czerwonego ma numer 37. Vorsetzer w systemie 2.0, gdy obraca obraz z mapy bitowej tam i z powrotem, zerka w dziurę, czekając na mignięcie numeru 37. Innymi słowy, pytam, czy którykolwiek z demonów pikselowych chce mu powiedzieć: „Mam tu kolor numer 37". Cała czerwień zniknęła – pozostały tylko liczby. W końcu cała praca w systemie CAD Dla Niewidomych musi zostać wykonana przez operacje arytmetyczne na ciągach bitów, tak jak to obserwowaliśmy na najniższym poziomie Shakeya w rozdziale 4. I nieważne, jak quasi-obrazkowe czy wyobrażeniowe są procesy prowadzące do werbalnej odpowiedzi Vorsetzera na pytania, nie będą wygenerowane w wewnętrznym miejscu, gdzie zagubione właściwości (właściwości jedynie „mówiły o" w mapie bitowej) są w jakiś sposób *przywracane*, aby mogły zostać *ocenione* przez sędziego tworzącego odpowiedzi.

Ludzie nie są systemami CAD Dla Niewidomych. Fakt, że system ten może manipulować swoimi „wyobrażeniami umysłowymi" i analizować je bez teatru kartezjańskiego, nie udowadnia sam w sobie, że w mózgu człowieka nie ma takiego teatrzyku, ale pokazuje, iż nie musimy się odnosić do niego, wyjaśniając ludzki talent do rozwiązywania problemów „oczyma wyobraźni". Istnieją procesy umysłowe silnie *analogiczne* do obserwacji, lecz gdy zredukujemy metaforę kineskopu Kosslyna do samej jej istoty, usuniemy właśnie te cechy, które nazwalibyśmy „teatrem kartezjańskim". Nie potrzeba żadnego czasu ani miejsca, gdzie „to wszystko się łączy" ku pożytkowi jednego, zunifikowanego obserwatora; rozróżnienia mogą się dokonać w sposób rozproszony, asynchroniczny i wielopoziomowy.

2. Słowa, obrazy i myśli

Prawdziwie „twórczy" aspekt języka leży nie w jego „nieskończonej zdolności do generowania", ale w cyklach realizacji i rozumienia, zapośredniczonych przez umysł zdolny do refleksji nad wieloma znaczeniami, które można przypisać zdaniu, znaczeniami, które nie musiały być obecne w myśli tworzącej to zdanie, ale które stały się dostępne poprzez samozrozumienie (lub głęboką interpretację zdania innej osoby) i mogą prowadzić do nowych myśli, wyrażanych i reinterpretowanych, i tak dalej, w nieskończoność.

H. Stephen Straight, 1976, s. 540

Brytyjskiego ekonoma Johna Maynarda Keynesa zapytano kiedyś, czy myśli słowami, czy obrazami. „Myślę myślami" – odpowiedział. Słusznie oparł się sugestii, że „rzeczy, którymi myślimy" są słowami lub obrazami, gdyż – jak widzieliśmy – „obrazy umysłowe" nie są

dokładnie takie jak obrazy w głowie, a myślenie „werbalne" nie jest *dokładnie* takie jak mówienie do siebie. Jednak powiedzenie, że ktoś myśli myślami, nie jest lepsze. Jedynie oddala pytanie, ponieważ myśl jest właśnie *tym, cokolwiek się dzieje, gdy myślimy* – a co to jest dokładnie, to już nie takie jasne.

Skoro przyjrzeliśmy się szkicom rodzajów drugoplanowej maszynerii odpowiedzialnej przyczynowo za szczegóły naszych światów heterofenomenologicznych, możemy *rozpocząć* opisywanie fenomenologii myślenia, wyjaśniając nie tylko ograniczenia i warunki fenomenologii „wizualnej" i „werbalnej", ale szukając jeszcze innych jej rodzajów, które nie wymykają się tej dychotomii.

Jednym z moich ulubionych ćwiczeń z zakresu heterofenomenologii jest książka Vladimira Nabokova *Obrona Łużyna* o mistrzu Łużynie, szachowym geniuszu, który ma załamanie nerwowe podczas kulminacyjnej rozgrywki. Widzimy trzy etapy rozwoju jego świadomości: jego chłopięcy umysł (zanim odkrył szachy w wieku około dziesięciu lat), umysł przesycony szachami (do momentu załamania nerwowego) oraz przykrą pozostałość dwóch pierwszych etapów po załamaniu, gdy uwięziony przez swoją żonę w świecie bez szachów – bez rozmowy o szachach, bez gry, bez książek o szachach – jego umysł powraca do pewnego rodzaju rozpuszczonej, infantylnej paranoi, rozjaśnianej wykradzionymi momentami szachowymi – rozjaśniającymi ukradkowymi atakami na diagramy szachowe w gazetach – jednak ostatecznie poddaje się obsesjom szachowym, które kończą się jego samomatem. Dowiadujemy się, że Łużyn ma umysł tak przesiąknięty szachami, iż widzi całe swoje życie w ten sposób. Oto przedziwne zaloty kobiety, która później zostanie jego żoną:

Łużyn, zacząwszy nagle mówić o tym, że niegdyś, jako młody chłopiec, mieszkał w tym hotelu, podjął ostrożnymi posunięciami, których sens uświadamiał sobie bardzo niejasno, swoiste wyznanie miłosne.

– Niech pan opowie coś jeszcze – powtórzyła, choć spostrzegła, jak chmurnie i ze znudzeniem zamilkł.

Siedział, opierając się na laseczce, i myślał o tym, że lipą stojącą na oświetlonym zboczu można by ruchem skoczka wziąć tamten oto słup telegraficzny, i zarazem usiłował sobie przypomnieć, o czym mówił przed chwilą. [...]

Przywierając ramieniem do jego piersi, próbowała ostrożnie unieść wyżej palcem jego powieki, i od lekkiego nacisku na gałkę oczną skakało dziwne, czarne światełko, skakało niczym jego czarny skoczek, który po prostu brał pionka, jeśli go Turati wysuwał w siódmym posunięciu, jak to zrobił podczas ostatniego spotkania. [Nabokov 1930/2005, s. 88, 104]

A oto spojrzenie na stan jego umysłu po załamaniu nerwowym:

Znalazł się w pełnym dymu pomieszczeniu, gdzie siedziały hałaśliwe widma. W każdym rogu dojrzewał atak – i potrącając stoliki, wiaderko, skąd sterczał szklany pion o złocistym gardle, bęben, w który bił, wygiąwszy się, grzywiasty szachowy koń, dobrnął do szklistego, cicho wirującego blasku [...]. [Nabokov 1930/2005, s. 125]

Te motywy są „obrazami" pod wieloma względami, gdyż szachy to gra przestrzenna i nawet tożsamość figur jest standardowo określona przez ich kształty, lecz władza szachów nad umysłem Łużyna nie wyczerpuje się w ich wizualnych czy przestrzennych własnościach – wszystko, co może zostać uchwycone na zdjęciach lub filmach szachownicy, jej figury w ruchu. Te cechy wizualne stanowią tak naprawdę jedynie najpłytszy aspekt jego wyobraźni. O wiele silniejsza jest *dyscyplina* zasad i taktyki gry; to abstrakcyjna struktura szachów, z którą tak obsesyjnie się zapoznał, a jego nawyki związane z badaniem tej struktury prowadzą jego umysł od „myśli" do „myśli".

[...] naraz, połapawszy się, stwierdzał smętnie, że znowu nie dopatrzył i w jego życiu

dokonał się subtelny ruch, bezlitośnie kontynuujący feralną kombinację. Postanowił więc podwoić czujność, obserwować każdą sekundę życia, wszędzie bowiem mógł się czaić podstęp. Najbardziej przy tym dręczyła go niemożność obmyślenia rozumnej obrony, bowiem cel przeciwnika był jeszcze niewiadomy. [Nabokov 1930/2005, s. 203]

Ucząc się jeździć na rowerze lub prowadzić samochód, odkrywamy nowe struktury możliwości działania, z ich ograniczeniami, znakami, rutynami, perspektywami, rodzaj abstrakcyjnego labiryntu zachowań, w którym szybko się odnajdujemy. Wkrótce stają się „drugą naturą". Sprawnie włączamy strukturę tego zewnętrznego zjawiska w swoją własną strukturę kontrolną. W tym procesie możemy mieć okresy obsesyjnej eksploatacji, kiedy to nie jesteśmy w stanie oderwać się od nowo poznanych ruchów. Pamiętam krótki okres mojej manii brydżowej, gdy mając kilkanaście lat, obsesyjnie i bezsensownie śniłem o brydżu. Wykonywałem ten sam ruch setki razy albo śniłem o „licytowaniu" podczas rozmów z moimi nauczycielami i kolegami z klasy. Moje omamy hipnagogiczne (owe dosyć halucynacyjne momenty, które nawiedzają nas, gdy zasypiamy lub gdy właśnie się budzimy) były pełne problemów typu „jaka jest prawidłowa reakcja na zalicytowanie trzech książek – cztery noże czy cztery widelce?".

Napotykając nową, abstrakcyjną strukturę w świecie – zapis muzyczny, język programowania, powszechne prawo, pierwszą ligę baseballową – dość częste jest pokonywanie tej ścieżki tam i nazad, co tworzy myślowe koleiny – przez próbę ich prawdziwego zrozumienia i wyczucia. Łużyn jest przypadkiem ekstremalnym; ma tylko jedną strukturę, którą może się bawić, i używa jej do wszystkiego. I w końcu zaczyna ona dominować nad innymi strukturami w jego umyśle, kierując sekwencjami jego myśli niemal tak sztywno, jak ciągi instrukcji w programie maszyny von Neumanna.

Pomyślmy o wszystkich strukturach, których uczą w szkole i wszędzie indziej: odczytywanie zegara, arytmetyka, pieniądze, trasy autobusów, korzystanie z telefonu. Ale ze wszystkich struktur, z którymi zapoznajemy się w ciągu życia, z pewnością najbardziej wszechobecnym i potężnym źródłem dyscypliny w naszych umysłach jest nasz język ojczysty. (Zwykle dostrzegamy najwięcej, przyglądając się kontrastom; Oliver Sacks w *Zobaczyć głos* [1989/1998] doskonale zwraca uwagę na bogactwa, których dostarcza umysłowi język, pokazując okrutne zubożenie umysłu niesłyszącego dziecka, jeśli jest ono pozbawione wczesnego dostępu do języka *naturalnego* – języka migowego). W rozdziale 8 widzieliśmy, jak bardzo samo słownictwo, które mamy do użytku, wpływa nie tylko na sposób, w jaki mówimy do innych, ale również na to, jak mówimy do siebie. Poza czynnikiem *leksykalnym* istnieje czynnik *gramatyczny*. Jak wskazuje Levelt (1989, sekcja 3.6), obligatoryjne struktury zdań w naszych językach są jak przewodnicy przypominający o tym, by sprawdzić to i odnieść się do tamtego, *wymagający* od nas organizowania faktów w pewien konkretny sposób. Część tej struktury może rzeczywiście być wrodzona, jak uważa Chomsky i inni, ale nie ma tak naprawdę znaczenia, gdzie leży granica dzieląca struktury ulokowane w mózgu genetycznie i te, które docierają do niego jako memy. Te struktury, prawdziwe czy wirtualne, tworzą szlaki, po których mogą następnie „podróżować" myśli.

Język zakaża i odmienia nasze myśli na każdym poziomie. Słowa, które znamy, są katalizatorami mogącymi wytrącić utrwalone treści, gdy jedna część mózgu próbuje porozumieć się z drugą. Struktury gramatyczne narzucają dyscyplinę naszym myślowym nawykom, kształtując sposoby, w jakie badamy nasze własne „bazy danych", próbując, jak hodowca ptaków u Platona, przywołać odpowiednie ptaki. Struktury historii, których się uczymy, zapewniają wskazówki na innym poziomie, zachęcając nas do zadawania samym sobie pytań, które mogą z największym prawdopodobieństwem przydać się nam w bieżących okolicznościach.

Nic z tego nie ma sensu, dopóki nadal ujmujemy umysł jako idealnie racjonalny

i doskonale dla siebie przejrzysty czy jednolity. Co miałoby mieć na celu *mówienie do siebie*, jeśli wiadomo już, co masz zamiar powiedzieć? Gdy jednak dostrzeżemy możliwość częściowego zrozumienia, niedoskonałej racjonalności, problematycznej komunikacji między różnymi obszarami, możemy zobaczyć, jak ogromne siły wyzwalane przez język w mózgu mogą być wykorzystywane w różnych formach podnoszenia siebie samego na wyższy poziom (*bootstrapping*), niektórych dobroczynnych, a innych – złośliwych.

Oto przykład:

Jesteś wspaniały!

A oto kolejny przykład:

Jesteś żałosny!

Wiesz, co oznaczają te zdania. Wiesz również, że właśnie przywołałem je znikąd, jako pomoc w wyjaśnieniu pewnej kwestii filozoficznej, oraz że nie są to zamierzone wypowiedzi żadnej osoby. Z pewnością ani ci nie schlebiam, ani cię nie obrażam, a nie ma tu nikogo innego. Ale czy można schlebiać sobie lub obrażać siebie, korzystając z jednego z moich zdań i wypowiadając je wielokrotnie do siebie „z naciskiem"? Spróbuj, jeśli masz odwagę. Coś się dzieje. Przez pierwszą minutę sobie nie wierzysz (mówisz do siebie), lecz stwierdzasz, że powtarzanie sobie tych słów rozbudza pewne reakcje, być może nawet lekkie zaróżowienie uszu, łącznie z reakcjami, odpowiedziami, sprostowaniami, obrazami, wspomnieniami, projektami. Te reakcje mogą oczywiście mieć różne rezultaty. Dale Carnegie miał rację co do mocy pozytywnego myślenia, ale jak z większością technik, łatwiej je stworzyć, niż kontrolować. Gdy mówisz do siebie, nie musisz sobie wierzyć, aby pojawiły się reakcje. Z pewnością one nastąpią i z pewnością będą miały związek w taki czy inny sposób ze znaczeniem słów, którymi się stymulujesz. Kiedy już rozpoczną się reakcje, mogą one poprowadzić twój umysł w miejsca, w których okaże się, że jednak sobie wierzysz – uważaj zatem, co do siebie mówisz.

Filozof Justin Leiber podsumowuje rolę języka w kształtowaniu naszego życia umysłowego:

Patrząc na siebie z punktu widzenia komputera, nie możemy uniknąć dostrzeżenia, że język naturalny jest naszym najważniejszym „językiem programowania". Oznacza to, że ogromna część naszej wiedzy i czynności jest przez nas najlepiej komunikowana i rozumiana w naszym języku naturalnym. [...] Można by powiedzieć, że język naturalny był naszym pierwszym wielkim, oryginalnym wytworem i, co sobie coraz bardziej uświadamiamy, skoro języki to maszyny, to język naturalny wraz z mózgiem, który nim steruje, był wynalazkiem na miarę komputera uniwersalnego. Można by to powiedzieć, gdyby nie ukradkowe podejrzenie, że język nie jest czymś, co wynaleźliśmy, ale czymś, czym się staliśmy, nie czymś, co skonstruowaliśmy, ale czymś, w czym stworzyliśmy i odtworzyliśmy siebie samych. [Leiber, s. 8]

Hipoteza mówiąca, że język odgrywa kluczową rolę w myśleniu, na pierwszy rzut oka może się wydawać wersją często dyskutowanej hipotezy mówiącej, że istnieje *język myśli*, jeden nośnik, w którym przebiega całe poznanie (Fodor 1975). Jest jednak między nimi istotna różnica. Leiber trafnie nazywa język naturalny „językiem programowania mózgu", ale możemy odróżnić języki programowania wyższego poziomu (takie jak Lisp, Prolog czy Pascal) od podstawowych „języków maszynowych" lub odrobinę mniej elementarnych „asemblerów", z których owe języki programowania wyższego poziomu są budowane. Języki wyższego poziomu to maszyny wirtualne i tworzą one (tymczasowe) struktury w komputerze, wyposażające go w konkretny wzorzec silnych i słabych stron. Cena, którą płacimy za to, że pewne rzeczy „łatwo powiedzieć", jest taka, że inne powiedzieć „trudno" czy nawet nie sposób. Taka maszyna wirtualna może nadawać strukturę tylko części kompetencji komputera, pozostawiając inne części podstawowej

maszynerii nietknięte. Pamiętając o tym rozróżnieniu, przekonująca jest teza, że szczegóły języka naturalnego – słownictwo i gramatyka angielskiego, chińskiego czy hiszpańskiego – krępują mózg podobnie do języka programowania wyższego poziomu. Nie ma to jednak nic wspólnego z wyrażaniem wątpliwej hipotezy, że taki język naturalny zapewnia strukturę *do najniższego poziomu organizacji*. I rzeczywiście, Fodor i inni broniący idei języka myśli zwykle twierdzą, że *nie* mówią o poziomie, na którym języki *ludzkie* urzeczywistniają takie skrępowania. Mówią o głębszym, mniej dostępnym poziomie reprezentacji. Fodor zdobył się kiedyś na zabawne wyznanie: przyznał, że gdy bardzo ciężko myślał, jedynym rodzajem elementów językowych, których był świadom, były urywki typu: „No dalej, Jerry, dasz radę!". Mogły to być jego „myśli" i zobaczyliśmy właśnie, jak w rzeczywistości mogą one odgrywać istotną rolę w rozwiązywaniu problemów, z którymi musi się zmagać, lecz nie do końca są czymś, z czego możemy stworzyć percepcyjne wnioski, hipotezy do przetestowania oraz inne postulowane operacje na podstawowym poziomie języka myśli. Keynes miał rację, opierając się wyborowi między słowami a obrazami; nośniki używane przez mózg są jedynie odrobinę analogiczne do reprezentacyjnych nośników życia publicznego.

3. Relacjonowanie i wyrażanie

Pomału osłabiamy ideę teatru kartezjańskiego. Naszkicowaliśmy konkurencyjną wobec centralnego nadawacza sensu hipotezę w rozdziale 8, a przed chwilą widzieliśmy, jak oprzeć się urokowi wewnętrznego kineskopu. Obawiam się, że to jedynie boczny cios; teatr kartezjański nadal ma się dobrze, nadal nieustępliwie wpływa na naszą wyobraźnię. Czas na zmianę taktyki i atak od wewnątrz, wybuch w teatrze kartezjańskim przez pokazanie jego niespójności na jego własnych zasadach. Zobaczmy, co się stanie, gdy poddamy się tradycji, po drodze akceptując zasady codziennej „psychologii potocznej", uznając je za poprawne. Możemy zacząć od ponownego przyjrzenia się niektórym przekonującym twierdzeniom Ottona z początku rozdziału 8:

Gdy *ja* mówię – kontynuuje Otto – mam na myśli to, co mówię. Moje świadome życie jest prywatne, ale mogę postanowić, że odkryję przed tobą pewne jego aspekty. Mogę zdecydować, że powiem ci o pewnych rzeczach dotyczących mojego obecnego lub wcześniejszego doświadczenia. Gdy to robię, formułuję zdania, które ostrożnie dostosowuję do materiału, z którego chcę zdać sprawę. Mogę poruszać się tam i z powrotem między przeżyciami a potencjalnym sprawozdaniem, konfrontując słowa z przeżyciem, upewniając się, że odnalazłem odpowiednią formę wyrażającą dokładnie to, co chcę. Czy wino ma w smaku aromat *grejpfruta*, czy może bardziej przypomina mi *jagody*? Czy byłoby lepiej powiedzieć, że wyższy ton brzmiał *głośniej*, czy po prostu wydawał on się *wyraźniejszy* lub *bardziej natężony*? Odnoszę się do mojego szczególnego, świadomego przeżycia i dokonuję osądu, które słowa najlepiej oddają jego charakter. Gdy jestem usatysfakcjonowany precyzyjnym wysłowieniem, wyrażam je. Z mojego introspekcyjnego raportu możesz dowiedzieć się o pewnej cesze mojego świadomego przeżycia.

Częściowo ta teza świetnie pasuje do proponowanego przez nas modelu realizacji języka z rozdziału 8. Proces ciągłego dopasowywania słów do treści przeżycia może być dostrzeżony w pandemonium łączącym demony słów z demonami treści. Brakuje oczywiście wewnętrznego Ja, którego oceny kierują tymi połączeniami. Lecz chociaż Otto rzeczywiście mówi o tym, co „*ja* wybieram" i co „*ja* osądzam", introspekcja tego nie potwierdza.

Mamy niewielki dostęp do procesów, dzięki którym słowa „pojawiają" nam się do powiedzenia, nawet wówczas, gdy mówimy celowo, próbując nasze wypowiedzi po cichu, zanim wypowiemy je na głos. Potencjalne wypowiedzi do powiedzenia po prostu wyskakują nie

wiadomo skąd. Okazuje się, że już je wypowiadamy lub że jakoś je sprawdzamy, czasem je odrzucając, a czasem trochę je redagując przed wypowiedzeniem, ale nawet te sporadyczne kroki pośrednie nie dają nam żadnych wskazówek co do tego, jak to robimy. Po prostu akceptujemy jakieś słowo lub je odrzucamy. Jeśli mamy racje osądów, to rzadko są rozważane przed wykonaniem, a jedynie stają się oczywiste po fakcie. („Miałam zamiar użyć słowa *jałowy*, ale się powstrzymałam, bo brzmiałoby to zbyt pretensjonalnie"). Zatem nie mamy uprzywilejowanego dostępu do procesów zachodzących w nas, gdy myśli stają się mową. O ile wiemy, *mogą* być tworzone przez pandemonium.

Mimo wszystko – kontynuuje Otto – model pandemonium pomija poziom czy etap tego procesu. Twojemu modelowi brakuje nie rzutowania do „przestrzeni fenomenalnej" teatru kartezjańskiego – co za absurdalny pomysł! – ale dodatkowego poziomu artykulacji w psychologii mówiącego. Nie wystarczy, że słowa się ze sobą złączą w jakimś wewnętrznym tańcu godowym, a następnie zostaną wypowiedziane. Jeśli mają być *raportami* z czyichś stanów umysłowych, muszą w jakiś sposób opierać się na akcie wewnętrznego *zrozumienia*. Model pandemonium pomija stan świadomości (*awareness*) mówiącego, który to stan kieruje mową.

Bez względu na to, czy Otto ma rację, czy nie, z pewnością wyraża powszechną mądrość: właśnie w taki sposób zwykle postrzegamy umiejętność mówienia ludziom o naszych świadomych stanach. W serii niedawnych artykułów filozof David Rosenthal (1986, 1989, 1990a, 1990b) zanalizował tę codzienną koncepcję świadomości i jej związek z naszymi pojęciami *raportowania* i *wyrażania*. Pokazuje pewne strukturalne cechy, które możemy dobrze wykorzystać. Po pierwsze, możemy wykorzystać jego analizę, aby zobaczyć od wewnątrz, jak wygląda standardowe postrzeganie i dlaczego jest tak nieodparte. Po drugie, możemy pokazać, jak dyskredytuje ideę zombi – bez pomocy z zewnątrz. Po trzecie, możemy obrócić standardowy pogląd przeciwko niemu samemu i wykorzystać napotkane trudności, aby nakreślić lepszy obraz, zachowujący ziarno prawdy w poglądzie tradycyjnym, ale odrzucający kartezjańskie ramy pojęciowe.

Co się dzieje, gdy mówimy? Sednem powszechnego mniemania na ten temat jest pewien truizm: Mówimy, co myślimy, o ile nie kłamiemy i nie jesteśmy nieszczerzy. Dokładniej można powiedzieć, że wyrażamy jedno z naszych przekonań lub myśli. Załóżmy na przykład, że widzisz kota nerwowo czekającego obok lodówki i mówisz: „Kot chce kolację". Wyraża to twoje przekonanie, że kot chce kolację. Przez *wyrażenie* swojego przekonania, *zdajesz sprawę* z tego, co uważasz za fakt dotyczący kota. W tym przypadku składasz raport o pragnieniu kota chcącego jeść. Warto zauważyć, że nie *zdajesz sprawy* ze swojego przekonania ani też nie *wyrażasz* pragnienia kota. Kot wyraża swoje pragnienie, stojąc niespokojnie obok lodówki, a ty, zauważywszy to, używasz tego jako podstawy – dowodu – swojego raportu. Istnieje wiele sposobów wyrażenia stanu umysłowego (takiego jak pragnienie), ale tylko jeden sposób zdania z niego sprawy, czyli wykonanie aktu mowy (ustnego, pisemnego czy inaczej przekazanego).

Ryc. 10.7

Jednym z najbardziej interesujących sposobów *wyrażania* stanu umysłowego jest *zdanie sprawy* z innego stanu umysłowego. W powyższym przykładzie relacjonujesz pragnienie kota, tym samym wyrażając swoje własne przekonanie o pragnieniu kota. Twoje zachowanie świadczy nie tylko o tym, że kot ma pragnienie, ale również o tym, że masz przekonanie, iż kot ma pragnienie. Możesz jednak dać świadectwo swojego przekonania w inny sposób – być może przez ciche powstanie z krzesła i przygotowanie kolacji dla kota. *Wyraziłoby* to identyczne przekonanie bez żadnego *relacjonowania*. Można także po prostu usiąść na krześle i przewrócić oczami, *nieumyślnie* wyrażając swoją irytację na pragnienie kota dokładnie wtedy, gdy wygodnie usiadło się na krześle. Wyrażanie stanu umysłowego, umyślne bądź nie, jest po prostu robieniem

czegoś, co stanowi dobre świadectwo tego stanu lub ujawnia go innemu obserwatorowi – osobie, którą nazwiemy „czytającą w myślach". Inaczej jest z relacjonowaniem stanów umysłowych, co jest czynnością bardziej wyszukaną, zawsze intencjonalną i językową.

Oto zatem ważna wskazówka dotycząca źródła modelu teatru kartezjańskiego: codzienna psychologia potoczna traktuje relacjonowanie stanów umysłowych jak relacjonowanie zdarzeń ze świata zewnętrznego. Twoja relacja, że kot chce kolację, opiera się na obserwacji kota. Relacja wyraża przekonanie, że kot chce jeść, przekonanie dotyczące pragnienia kota. Nazwijmy przekonania o przekonaniach, pragnienia o pragnieniach, przekonania o pragnieniach, nadzieje o obawach itp. „stanami umysłowymi *drugiego rzędu*". I jeśli (1) *jestem przekonany*, że ty (2) uważasz, że ja (3) *chcę* napić się kawy, to moje przekonanie jest przekonaniem trzeciego rzędu. (O ważności stanów umysłowych wyższego rzędu w teoriach umysłu przeczytasz w mojej książce *Intentional Stance* [Dennett 1987a]), Nie ma żadnych wątpliwości, że istnieją wyraźne różnice odpowiadające tym codziennym rozróżnieniom, gdy są stosowane niezwrotnie – gdy *x* uważa, że *y* jest w jakimś stanie umysłowym, a $x \neq y$. Jest ogromna różnica między przypadkiem, w którym kot chce zostać nakarmiony i ty o tym wiesz, a przypadkiem, w którym kot chce zostać nakarmiony i ty o tym nie wiesz. Ale co z przypadkami zwrotnymi, gdzie $x = y$? Psychologia potoczna traktuje te przypadki dokładnie tak samo.

Załóżmy, że relacjonuję, że *ja* chcę jeść. W modelu standardowym muszę wyrażać *przekonania drugiego rzędu* o *moim* pragnieniu. Gdy relacjonuję moje pragnienie, wyrażam pogląd drugiego rzędu – mój pogląd o moim pragnieniu. A co jeśli *zrelacjonuję* to przekonanie drugiego rzędu, mówiąc: „Uważam, że chcę być nakarmiony"? Ta relacja musi wyrażać przekonanie *trzeciego* rzędu – moje przekonanie, że rzeczywiście chcę zostać nakarmiony. I tak dalej. Nasze codzienne pojęcia tego, na czym polega mówienie, naprawdę tworzą w ten sposób mnóstwo przypuszczalnie odmiennych stanów umysłowych: moje pragnienie jest różne od mojego przekonania, że mam pragnienie, co z kolei różni się od mojego przekonania, że jestem przekonany, że mam pragnienie itd.

Psychologia potoczna czyni dalsze rozróżnienia. Jak zwraca uwagę Rosenthal (wraz z wieloma innymi), odróżnia ona *przekonania*, które są podstawą stanów dyspozycyjnych, od *myśli*, które są stanami bieżącymi lub epizodycznymi – zdarzeniami przemijającymi. Twoje *przekonanie o tym, że psy to zwierzęta* utrzymuje się nieprzerwanie jako stan twojego umysłu od lat, ale zwracając na to uwagę właśnie teraz, wywołałem w tobie *myśl* – myśl, że psy to zwierzęta, epizod, który bez wątpienia nie zdarzyłby się w tobie właśnie teraz bez mojego udziału.

Wynika z tego oczywiście, że mogą istnieć myśli rzędu pierwszego, drugiego oraz myśli wyższego rzędu – myśli o myślach (o myślach...). I tu pojawia się kluczowy krok: Gdy wyrażam przekonanie – na przykład przekonanie o tym, że chcę jeść – nie wyrażam przekonania wyższego rzędu bezpośrednio; moje przekonanie prowadzi do epizodycznej myśli, *myśli* wyższego rzędu o tym, że chcę jeść, i wyrażam *tę myśl* (jeśli tak postanowię). Wszystko to jest elementem, jak twierdzi Rosenthal, zdroworozsądkowego modelu *mówienia tego, co myślisz*.

Cechą swoistą stanów ludzkiej świadomości jest to, że mogą być relacjonowane (wykluczając na przykład afazję, paraliż czy zakneblowanie ust), więc wynika z tego, według analizy Rosenthala, że „świadome stany muszą występować w towarzystwie odpowiednich myśli wyższego rzędu, a nieświadome stany umysłowe nie mogą mieć takich towarzyszy" (1990b, s. 16). Wspomniana myśl wyższego rzędu musi oczywiście dotyczyć stanu, któremu towarzyszy; musi być myślą o tym, że jest się w stanie niższego rzędu (lub właśnie się w nim było – czas szybko mija). Wygląda to, jakby miał się stąd wyłaniać nieskończony regres stanów świadomych czy myśli wyższego rzędu, ale Rosenthal twierdzi, że psychologia potoczna pozwala na

niesamowitą inwersję: *myśl drugiego rzędu nie musi sama w sobie być świadoma, aby jej pierwszorzędowy przedmiot był świadomy*. Możesz *wyrazić* myśl bez jej świadomości, więc możesz wyrazić myśl *drugiego rzędu* bez *jej* świadomości – musisz być świadomy jedynie jej przedmiotu, pierwszorzędowej myśli, którą *relacjonujesz*.

Z początku może się to wydawać zaskakujące, ale po zastanowieniu można zauważyć, że to znany fakt widziany z nowej perspektywy: nie odnosisz się do wyrażanej myśli, ale do przedmiotu/przedmiotów, których *dotyczy* ta myśl. Rosenthal następnie twierdzi, że choć niektóre myśli drugiego rzędu *są* świadome – z racji trzeciorzędowych myśli o nich – są one dość rzadkie. Są zdecydowanie introspekcyjne, a zrelacjonowalibyśmy je (nawet sobie samym) tylko wówczas, gdybyśmy byli w stanie supersamoświadomości. Jeśli powiem ci „boli mnie", relacjonuję świadomy stan, mój ból, oraz wyrażam przekonanie drugiego stopnia – moje przekonanie o tym, że mnie boli. Jeśli, naprawdę filozoficznie, powiem „myślę [lub jestem pewien, lub mam przekonanie], że mnie boli", w ten sposób relacjonuję myśl drugiego rzędu, wyrażając myśl trzeciego rzędu. Zwykle jednak nie miałbym takiej trzeciorzędowej myśli i wówczas nie byłbym świadom takiej myśli drugiego rzędu; wyraziłbym ją, mówiąc „boli mnie", ale nie byłbym tego w normalny sposób świadomy.

Idea nieświadomych myśli wyższego rzędu z początku może się wydawać oburzająca lub paradoksalna, lecz kategoria epizodów, o których mowa, nie jest kontrowersyjna, nawet jeśli pojęcie „myśl" zwykle nie jest używane na ich oznaczenie. Rosenthal wykorzystuje „myśl" jako pojęcie techniczne – z grubsza idąc śladami Kartezjusza – aby wyrazić za jego pomocą *wszystkie* stany epizodyczne o jakiejś treści, a nie tylko epizody, które zwykle nazwalibyśmy „myślami". W ten sposób ukłucie bólu lub ujrzenie pończochy według Kartezjusza i Rosenthala liczyły się jako myśl. Jednak w przeciwieństwie do Kartezjusza Rosenthal twierdzi, że istnieją nieświadome myśli.

Nieświadome myśli to na przykład nieświadome zdarzenia percepcyjne, epizodyczne aktywacje przekonań, które zachodzą w sposób naturalny – które *muszą* zachodzić – podczas zwykłego kontrolowania zachowania. Przypuśćmy, że przewrócisz filiżankę z kawą na biurku. Jak błyskawica zrywasz się z krzesła, w ostatniej chwili unikając wylania kapiącej ze stołu kawy na siebie. Nie masz świadomości myślenia, że blat biurka nie wchłonie kawy ani że kawa, płyn poddający się prawom grawitacji, zacznie skapywać z biurka, jednak takie nieświadome myśli musiały się pojawić – bo gdyby filiżanka była pełna soli albo gdyby biurko przykryte było ręcznikiem, nikt nie zerwałby się z krzesła. Ze wszystkich twoich przekonań – dotyczących kawy, demokracji, baseballu, ceny herbaty w Chinach – te oraz kilka innych były bezpośrednio związane z okolicznościami. Jeśli przytaczamy je, wyjaśniając, dlaczego zrywamy się z krzesła, musiały one przez moment być dostępne, aktywowane lub w inny sposób uzyskane, wpływając na zachowanie, ale oczywiście wydarzyło się to nieświadomie. Takie nieświadome epizody byłyby przykładami tego, co Rosenthal nazywa nieświadomymi myślami. (Natknęliśmy się już na nieświadome myśli we wcześniejszych przykładach: nieświadoma percepcja wibracji w palcach umożliwiająca świadome zidentyfikowanie faktury dotykanej patykiem, nieświadome przypomnienie sobie kobiety w okularach, które doprowadziło do błędnego przeżycia kobiety przebiegającej obok).

Rosenthal (1990b) wskazuje, że definiując świadomość w kategoriach nieświadomych stanów umysłu (towarzyszących myśli wyższego rzędu), odkrył *w psychologii* potocznej fundamenty pod niekolistą, pozbawioną tajemnicy teorię świadomości. Uważa, że stan świadomy od nieświadomego odróżnia nie jakaś niedająca się wyjaśnić własność wewnętrzna, ale prosta własność, którą jest posiadanie towarzyszącej myśli wyższego rzędu dotyczącej stanu, o którym mowa. (Harnad 1982 przedstawia podobną strategię, choć z pewnymi ciekawymi różnicami).

Psychologii potocznej rokuje to dobrze: nie jest oparta na tajemnicy; ma środki świetnie zanalizowane przez Rosenthala, pozwalające ująć cenioną kategorię, świadomość, w kategoriach podrzędnych i mniej problematycznych. Zyskiem z tej analizy jest to, że możemy jej użyć do obalenia pozornie ostrego rozróżnienia między istotami świadomymi a zombi.

4. Zombi, zimbo i iluzja użytkownika

Umysł to wzorzec postrzegany przez umysł. Być może jest to koliste, ale nie jest ani błędne, ani paradoksalne.

Doughlas Hofstadter, 1981, s. 200

Przypomnijmy sobie, że zombi filozofów pozornie dokonują aktów mowy, wydają się relacjonować swoje stany świadomości, wydają się dokonywać introspekcji. Lecz w rzeczywistości w ogóle nie są świadome, mimo że są nieodróżnialne od osoby świadomej, przynajmniej w swej najlepszej wersji. Mogą mieć stany wewnętrzne obdarzone funkcjonalną treścią (rodzajem treści, którą funkcjonaliści mogą przypisać wewnętrznej maszynerii robotów), ale są to stany nieświadome. Shakey – wyobrażony przez nas – jest typowym zombi. Kiedy „relacjonuje" stan wewnętrzny, nie relacjonuje świadomego stanu, ponieważ Shakey nie ma stanów świadomych, tylko nieświadomy stan, który jedynie powoduje, że przechodzi on w kolejny nieświadomy stan kierujący procesem generowania i wykonywania tak zwanych aktów mowy składających się z „prefabrykowanych" formułek. (Cały czas pozwalamy Ottonowi na takie twierdzenia).

Shakey nie podjął *najpierw* decyzji o tym, co zrelacjonować, po zaobserwowaniu czegoś, co odbywało się wewnątrz, a *następnie* wymyślił, jak to *wyrazić*; Shakey po prostu miał coś do powiedzenia. Shakey nie miał żadnego dostępu do tego, dlaczego chciał powiedzieć, że tworzył liniowe rysunki zgodnie z granicą między jasnymi i ciemnymi obszarami w jego obrazach umysłowych – po prostu był tak skonstruowany. Głównym twierdzeniem w rozdziale 8 było jednak to, że przeciwnie do tego, co na pierwszy rzut oka mogłoby się wydawać, to samo jest prawdą o nas. Nie mamy jakiegoś szczególnego dostępu do tego, dlaczego chcemy powiedzieć to, co stwierdzamy, że chcemy powiedzieć; po prostu jesteśmy tak skonstruowani. Jednak w przeciwieństwie do Shakeya bezustannie się przebudowujemy, odkrywając nowe rzeczy, które chcemy powiedzieć w wyniku zastanowienia nad tym, co właśnie się okazało, że chcemy powiedzieć, i tak dalej.

Ale czy bardziej zaawansowany Shakey również nie mógłby tego zrobić? Shakey był szczególnie topornym zombi, a teraz możemy sobie wyobrazić zombi bardziej realistycznego i złożonego, monitorującego swoje własne czynności, łącznie ze swoimi własnymi czynnościami wewnętrznymi, w nieskończonej, rosnącej spirali refleksyjności. Nazwę taką refleksyjną istotę „*zimbo*". Zimbo to zombi, który w wyniku monitorowania samego siebie ma wewnętrzne (ale nieświadome) stany informacyjne wyższego rzędu dotyczące innych stanów informacyjnych niższego rzędu. (W tym eksperymencie myślowym nie ma znaczenia, czy zimbo jest uważany za robota, czy za człowieka – albo Marsjanina). Ci, którzy uważają, że pojęcie zombi jest spójne, muszą z pewnością zaakceptować możliwość istnienia zimbo. Zimbo to po prostu zombi ze złożonymi zachowaniami, a to dzięki systemowi kontroli umożliwiającemu rekurencyjną autoreprezentację.

Zauważmy, jak zimbo dałby sobie radę z testem Turinga, zaproponowanym przez Alana Turinga w 1950 roku słynnym teście operacyjnym na myślenie w komputerach. Komputer potrafi myśleć, twierdził Turing, jeśli potrafi regularnie pokonać ludzkiego przeciwnika w „grze

w naśladowanie": dwóch zawodników jest ukrytych przed sędzią-człowiekiem, ale mogą się z nim komunikować, wpisując komunikat na terminalu komputerowym. Zawodnik-człowiek po prostu próbuje przekonać sędziego, że jest człowiekiem, a to samo robi zawodnik-komputer – próbuje przekonać sędziego, że jest człowiekiem. Jeśli sędzia nie potrafi wskazać na komputer, komputer zostaje uznany za myślący. Turing zaproponował ten test jako rozwiązanie kończące dyskusję; było dla niego jasne, iż ten test tak niesłychanie trudno zdać, że jakikolwiek komputer, który potrafiłby wygrać, powinien być przez wszystkich postrzegany jako *niesamowicie* dobrze myślący. Myślał, że ustawił poprzeczkę wystarczająco wysoko, aby usatysfakcjonować sceptyków. Przeliczył się. Wielu uważa, że „zdanie testu Turinga" nie jest wystarczającym dowodem inteligencji, a już na pewno nie świadomości. (Analizę silnych i słabych punktów testu Turinga oraz jego krytyków znajdziesz w: Hofstadter 1981b, Dennett 1985a i French 1991).

Szanse zimbo w teście Turinga powinny być takie same jak każdej świadomej osoby, ponieważ zawodnicy ujawniają sędziemu jedynie swoje zachowanie, a do tego tylko werbalne (pisemne). Załóżmy więc, że jesteś sędzią w teście Turinga, a (pozorne) akty mowy zimbo przekonały cię, że jest on świadomy. Te pozorne akty mowy nie powinny cię przekonać – *na mocy założenia*, gdyż to tylko zimbo, a zimbo nie są świadome. Czy zimbo powinno jednak przekonać samo siebie? Gdy zimbo tworzy relację, wyrażając swoje własne stany nieświadome drugiego rzędu, nie istnieje nic, co mogłoby zapobiec jego refleksji (nieświadomej) nad tym właśnie stanem rzeczy. Tak naprawdę, jeśli ma być przekonujący, będzie musiał być w stanie odpowiednio zareagować na swoje własne „asercje" komunikowane tobie (lub je zrozumieć).

Załóżmy na przykład, że zimbo jest bardziej zaawansowany od Shakeya, a ty jako sędzia właśnie pytasz go o rozwiązanie problemu oczyma wyobraźni, a następnie o wyjaśnienie, jak tego dokonał. Zastanawia się nad swoim własnym twierdzeniem, że właśnie rozwiązał ten problem, tworząc rysunek w wyobrażeniu umysłowym. „Wiedziałby", że jest to coś, co chciał powiedzieć, a gdyby zastanowił się dłużej, stwierdziłby, że „wie", iż nie wie, dlaczego chciał powiedzieć właśnie to. Im bardziej pytalibyśmy go o to, co wie i czego nie wie o tym, co robi, coraz głębiej by się nad tym zastanawiał. Wydaje się, że to, co właśnie sobie wyobraziliśmy, to nieświadoma istota, która mimo to może mieć myśli wyższego rzędu. Jednak według Rosenthala, gdy stanowi umysłowemu towarzyszy świadoma *lub* nieświadoma myśl wyższego rzędu o tym, że tę myśl ma, to tym samym gwarantuje to, że ten stan umysłowy jest stanem świadomym! Czy nasz eksperyment myślowy zdyskredytował analizę Rosenthala, czy zdyskredytował definicję zimbo?

Możemy łatwo dostrzec, że zimbo (nieświadomie) uważa, iż był w różnych stanach umysłowych – właśnie w tych stanach, z których może zdać relację, gdy go o to poprosimy. Myślałby, że jest świadomy, nawet jeśliby taki nie był! Każda jednostka będąca w stanie zdać test Turinga żyłaby w (błędnym?) poczuciu, że jest świadoma. Innymi słowy, byłaby ofiarą iluzji (zob. także Harnad 1982). Jakiej iluzji? Iluzji użytkownika, oczywiście. Byłaby „ofiarą" łagodnej iluzji użytkownika swojej własnej maszyny wirtualnej!

Czy nie jest to sztuczka z lustrami, jakiś niedozwolony rodzaj filozoficznego kuglarstwa? Jak może istnieć iluzja użytkownika bez teatru kartezjańskiego, w którym iluzja się pojawia? Wydaje się, że ostateczny cios zadadzą mi moje własne metafory. Problemem jest to, że iluzja użytkownika maszyny wirtualnej następuje przez prezentację materiału w jakiegoś rodzaju teatrze, w którym jest niezależna, zewnętrzna publiczność, użytkownik, dla którego całe przedstawienie jest wystawiane. W tym momencie korzystam z komputera, wpisując te słowa do „pliku" z dyskretną asystą edytora tekstu. Gdy wchodzę w interakcję z komputerem, mam ograniczony dostęp do tego, co się w nim dzieje. Dzięki systemom prezentacji stworzonym przez programistów mogę korzystać z rozbudowanej metafory audiowizualnej, interaktywnej sztuki

wystawianej na scenie klawiatury, myszki i ekranu. Ja, użytkownik, jestem poddany serii dobroczynnych iluzji: wydaje się, że jestem w stanie ruszyć kursorem (potężnym i widocznym sługą) dokładnie w miejsce komputera, gdzie trzymam mój plik, a gdy już zobaczę, że kursor „tam" dotarł, naciśnięciem klawisza otwieram plik, rozwijając jego długi zwój w oknie (na ekranie) zgodnie z moim poleceniem. Mogę sprawić, że różne rzeczy staną się w komputerze, wpisując do niego różne polecenia, wciskając odpowiednie klawisze, ale nie muszę znać szczegółów; zachowuję kontrolę, opierając się na moim rozumieniu szczegółowych metafor audiowizualnych zapewnionych przez iluzję użytkownika.

Większość użytkowników komputera jedynie dzięki tym metaforom może ocenić to, co dzieje się w jego wnętrzu. To między innymi dlatego maszyna wirtualna jest tak dobrą analogią dla świadomości, gdyż zawsze wydawało nam się, że nasz dostęp do tego, co dzieje się wewnątrz naszych mózgów, jest ograniczony; *my* nie musimy wiedzieć, w jaki sposób zakulisowa maszyneria naszych mózgów dokonuje magicznych sztuczek; *my* znamy jej działania tylko w sposób, w jaki je widzimy, gdy do nas docierają w interaktywnych metaforach fenomenologii. Ale jeśli skorzystamy z tej kuszącej analogii i podtrzymamy „oczywistą" separację pomiędzy prezentacją z jednej strony a oceną przedstawienia przez użytkownika z drugiej, wydaje się, że wylądujemy z powrotem w teatrze kartezjańskim. Jak może istnieć iluzja użytkownika bez tej separacji?

Nie może istnieć; użytkownik zapewniający perspektywę, z której maszyna wirtualna staje się „widoczna", musi być jakiegoś rodzaju zewnętrznym obserwatorem – *Vorsetzerem*. I można by pomyśleć, że idea takiego obserwatora musi być ideą obserwatora świadomego, ale widzieliśmy już, że wcale tak nie jest. Vorsetzer, który siadał przed systemem CAD w oryginalnym systemie CAD Dla Niewidomych 1.0, nie był świadomy, ale mimo wszystko miał tak ograniczony dostęp do wewnętrznych działań systemu CAD, jak każdy świadomy użytkownik. A w momencie gdy pozbędziemy się zbędnego ekranu z kamerą, prezentacja i ocena użytkownika wyparowują, zamienione – jak często w naszych relacjach – przez zastęp skromniejszych operacji. „Zewnętrzny obserwator" może być stopniowo włączony do systemu, pozostawiając za sobą zaledwie kilka śladów: fragmenty „interfejsu", którego różne formaty nadal ograniczają rodzaje pytań, na które można odpowiedzieć, i w ten sposób ograniczają wyrażalne treści[97]. Nie musi istnieć *jedno* miejsce, w którym odbywa się prezentacja[98]. A jak sugeruje nam analiza Rosenthala, nawet nasza zwykła koncepcja świadomości, jako zakotwiczona w rozsądnej intuicji psychologii potocznej, może tolerować nieświadomość stanów wyższego rzędu, których obecność w systemie wyjaśnia świadomość niektórych z jej stanów.

Czy proces nieświadomej refleksji jest więc ścieżką, na której zombi mógłby zamienić się w zimbo i *tym samym* stać się świadomy? Jeśli tak jest, to zombi muszą być jednak świadome. Wszystkie zombi potrafią wytworzyć przekonujące „akty mowy" (pamiętaj, są one nie do odróżnienia od naszych najlepszych przyjaciół), a ta umiejętność byłaby magiczna, gdyby struktury kontrolne lub procesy przyczynowe odpowiedzialne za to w mózgach zombi (lub w komputerze czy gdziekolwiek indziej) nie odzwierciedlały tych czynności i ich (pozornych czy funkcjonalnych) treści. Zombi mógłby zacząć swoją karierę w stanie niekomunikatywnym i bezrefleksyjnym, a wówczas naprawdę być zombi, ale gdy tylko zacząłby „komunikować się" z innymi i ze sobą, miałby stany takiego rodzaju, które według analizy Rosenthala wystarczają do świadomości.

Z drugiej strony, jeśli odrzucimy Rosenthala analizę świadomości w kategoriach myśli wyższego rzędu, wówczas zombi mogą żyć dalej i brać udział w kolejnych eksperymentach myślowych. Zaproponowałem tę przypowieść o zimbo ironicznie, gdyż nie uważam, że *ani* pojęcie zombi, *ani* kategorie myśli wyższego rzędu w psychologii potocznej nie mogą przetrwać

jako coś więcej niż odrzucona wiara w starych bogów. Rosenthal jednak wspaniale nam pomógł, odsłaniając logikę tych codziennych pojęć, i dzięki jaśniejszemu ich ujęciu możemy zobaczyć, co najlepiej by je zastąpiło.

5. Problemy z psychologią potoczną

Rosenthal stwierdza, że psychologia potoczna zakłada nieskończenie rozciągliwą hierarchię myśli wyższego rzędu, które są pojmowane jako wyraźne, niezależne, sensowne epizody wydarzające się w czasie rzeczywistym w umyśle. Czy taką wizję można potwierdzić empirycznie? Czy istnieją takie osobne stany i zdarzenia w mózgu? Przy odrobinie życzliwości odpowiedź musi brzmieć „tak". Z pewnością istnieją znane psychologiczne różnice, które mogą być – i zwykle są – opisywane w takich kategoriach.

Nagle Dorota uświadomiła sobie, że chciała wyjść – i że chciała to zrobić już od jakiegoś czasu.

Wydaje się tutaj, że Dorota uzyskała przekonanie drugiego rzędu – ponieważ miała myśl drugiego rzędu – o swoim pragnieniu jakiś czas po jego zaistnieniu. Istnieje wiele codziennych przypadków tego rodzaju: „I wówczas uświadomił sobie, że patrzył się wprost na brakującą spinkę do mankietu"; „Kocha ją – tylko jeszcze nie zdał sobie z tego sprawy". Trudno zaprzeczyć temu, że te zwykłe zdania są aluzją do prawdziwych zmian jednego „stanu umysłu" w drugi. A intuicyjnie, twierdzi Rosenthal, ta zmiana jest kwestią *uświadomienia* stanu pierwszego rzędu. Gdy Freud, opierając się na takich codziennych przypadkach, postuluje rozległą, ukrytą sferę nieświadomych stanów umysłowych, są one właśnie stanami, o których podmioty nie wiedzą, że się w nich znajdują. Ludzie ci są w stanach umysłowych, których jeszcze sobie nie uświadomili – poprzez myśli wyższego rzędu – że się w nich znajdują.

Ten sposób opisywania owych różnic jest znany, ale to, czy jest całkowicie klarowny, to inna kwestia. Są to wszystko przejścia w stan lepszego poinformowania (mówiąc tak neutralnie, jak się da), a bycie lepiej poinformowanym w ten sposób jest rzeczywiście koniecznym warunkiem *zrelacjonowania* (w przeciwieństwie do jedynie *wyrażania*) wcześniejszego „stanu umysłu". Nieostrożnie byłoby to ująć tak: *aby* zrelacjonować stan lub zdarzenie umysłowe, musisz mieć myśl wyższego rzędu, którą wyrażasz. Daje nam to obraz najpierw *obserwowania* (jakimś wewnętrznym organem zmysłowym) umysłowego stanu lub zdarzenia, *tym samym* wytwarzającego stan *przekonania*, którego początek wyznacza *myśl*, następnie wyrażona. Ten łańcuch przyczynowy, jak widzieliśmy, naśladuje łańcuch przyczynowy relacjonowania zwykłych zdarzeń zewnętrznych: najpierw obserwujesz wydarzenie za pomocą organów zmysłowych, które tworzą w tobie przekonanie, a następnie myśl, którą wyrażasz w swojej relacji.

Uważam, że ta przypuszczalna myśl wyższego rzędu jest „dodatkowym poziomem artykulacji", o którym Otto myślał, że może być wyróżniony w jego własnej psychice; właśnie tę myśl *wyrażają* słowa Ottona, gdy *relacjonuje* swoje własne świadome przeżycia. Jednak zgodnie z modelem realizacji mowy, o którym mówiliśmy w rozdziale 8, model Ottona na opak opisuje łańcuch przyczynowy. Nie jest tak, że *najpierw* wchodzimy w wyższego rzędu stan samoobserwacji, tworząc myśl wyższego rzędu, abyśmy następnie mogli zrelacjonować myśl niższego rzędu, wyrażając myśl wyższego rzędu. Stan drugiego rzędu (stan lepiej poinformowany) raczej zostaje *stworzony* właśnie przez proces tworzenia relacji. Nie rozumiemy naszego przeżycia *najpierw* w teatrze kartezjańskim, a *potem* na podstawie tej nabytej wiedzy nie mamy możliwości tworzenia relacji do wyrażenia; nasza *umiejętność powiedzenia*, jak to jest, *stanowi podstawę* naszych „przekonań wyższego rzędu"[99].

Z początku projekt pandemonium procesu aktu mowy wygląda na zły pomysł, ponieważ wydaje się pomijać centralnego obserwatora czy decydenta, którego myśl zostanie w końcu wyrażona. Jednak jest to siła, a nie słabość tego modelu. Pojawienie się wyrażenia jest właśnie tym, co tworzy lub ustala treść wyrażanej myśli wyższego rzędu. Nie musi istnieć *dodatkowa* epizodyczna „myśl". Stan wyższego rzędu dosłownie – przyczynowo – zależy od wyrażenia aktu mowy. Ale niekoniecznie od publicznego wyrażenia jawnego aktu mowy. W rozdziale 7 widzieliśmy, że potrzeba coraz lepszego komunikowania informacji przez organizm mogła doprowadzić do stworzenia nawyków automanipulacji, które mogły zająć miejsce ewolucyjnie bardziej pracochłonnego procesu tworzenia wewnętrznego oka, rzeczywistego wewnętrznego organu mogącego monitorować mózg. Stwierdziliśmy, że mózg może przejść w coś w rodzaju stanu przekonania wyższego rzędu tylko przez zaangażowanie się w proces podobny do relacjonowania stanów pierwszego rzędu samemu sobie.

Musimy zerwać z nawykiem przyjmowania coraz bardziej centralnych obserwatorów. Jako przejściową podporę możemy ponownie wyobrazić sobie ten proces nie jako wiedzę nabytą przez obserwację, ale jako model *pogłoski*. Uważam, że *p*, ponieważ zaufane źródło powiedziało mi, że *p*. Zaufane źródło, czyli kto? Ja – a w każdym razie jeden lub wielu z moich „agentów". Nie jest to myślenie zupełnie nie z tego świata; mówimy w końcu o *świadectwie* naszych zmysłów, co jest metaforą sugerującą, że nasze zmysły nie przynoszą dowodów rzeczowych do „sądu", aby je nam *pokazać*, a raczej *mówią* nam o pewnych rzeczach. Opierając się na tej metaforze (dopóki nie przyzwyczaimy się do złożonej hipotezy konkurencyjnej), możemy uciec się do dewizy:

Bez mówienia do samego siebie nie można wiedzieć, co się myśli.

Z kilku względów nie jest to jeszcze trafne ujęcie. Po pierwsze, jest różnica – którą dotychczas pomijałem – między bytem „mówiącym do siebie" a rozmaitością podsystemów „rozmawiających ze sobą". Odpowiednie przejście pomiędzy tymi dwiema ideami pojawi się w rozdziale 13 dotyczącym jaźni. Po drugie, jak już widzieliśmy, nacisk na wyrażenie językowe jest przesadą; istnieją inne strategie automanipulacji oraz autoekspresji, które nie są werbalne.

Może się komuś wydawać, że przedstawiam kiepską ofertę: porzucenie względnej zwięzłości i jasności standardowego modelu psychologii potocznej, wraz z jego hierarchią wewnętrznych obserwacji, na rzecz mglistej alternatywy, którą nadal ledwie co rozumiemy. Jednak klarowność tradycyjnego modelu jest iluzją z powodów, o których napomknąłem w rozdziale 5, gdzie badaliśmy osobliwy temat *rzeczywistego wydawania się*. Teraz możemy dokładniej zdiagnozować te problemy. Otto jest rzecznikiem psychologii potocznej, a jeśli pozwolimy mu kontynuować, wkrótce zacznie się plątać. Pogląd Ottona, który zawzięcie rozciąga kategorie psychologii potocznej „do oporu", tworzy eksplozję osobnych „stanów reprezentacyjnych", a relacje między nimi rodzą sztuczne dylematy. Otto kontynuuje:

Moje publiczne sprawozdanie ze stanu świadomego, jeśli postanowię je przedstawić, może zawierać w sobie błąd. Mogę się przejęzyczyć albo pomylić się co do tego, co jakieś słowo znaczy, i w ten sposób niechcący źle cię poinformować. Każdy taki błąd wyrazu, którego nie wychwyciłem, mógłby wywołać w tobie nieprawdziwe przekonanie o faktach – o tym, jak *naprawdę* jest ze mną. A sam fakt, że nie udało mi się wychwycić błędu, nie oznaczałby, że nie było błędu. *Z jednej strony* istnieje prawda dotycząca tego, jak to jest ze mną, ale *z drugiej strony* jest to, co ostatecznie mówię o tym, jak jest ze mną (jeśli postanowię to zrobić). Choć zwykle relacjonuję wysoce niezawodnie, zawsze pozostaje miejsce na wkradające się błędy.

Jest to jedna z tych sytuacji, gdzie można mówić tylko o dwóch stronach problemu. Gdyż, jak pokazał nam Rosenthal, oprócz faktu o tym, „jak to jest ze mną", oraz „tego, co ostatecznie mówię", istnieje też trzeci fakt: moje przekonanie o tym, jak to jest ze mną[100]. Jeśli bowiem

szczerze mówię to, co mówię, mam na myśli to, co mam na myśli, wyrażam jedno z moich przekonań – moje przekonanie o tym, jak to jest ze mną. Tak naprawdę wkracza tu czwarty fakt: moja epizodyczna myśl, że tak właśnie jest ze mną.

Czy mój pogląd na to, jak to jest ze mną, może być błędny? Czy może mogę *tylko myśleć*, że tak jest ze mną? Innymi słowy, czy może mi się jedynie *wydawać*, że to było moje obecne przeżycie? Otto chciał oddzielić jedną rzecz, ale teraz tych podziałów grozi nam więcej: między subiektywnym przeżyciem i przekonaniem na jego temat, między tym przekonaniem a epizodyczną myślą, którą powoduje on w drodze do wyrażenia werbalnego, oraz między tą myślą i jej ostatecznym wyrazem. I jak z mnożącymi się miotłami ucznia czarnoksiężnika, gdy zaakceptujemy poprzednie rozróżnienia, szybko pojawią się nowe. Załóżmy, że mam moje subiektywne przeżycia (to jedna kwestia) i stanowią one podstawę mojego przekonania, że je mam (to druga kwestia), co z kolei powoduje związaną z tym myśl (trzecia kwestia), która następnie pobudza mnie do intencji komunikacyjnej, aby tę myśl wyrazić (czwarta kwestia), co w końcu doprowadza do rzeczywistego wyrażenia (piąta kwestia). Czy nie ma miejsca na pomyłkę, która może się wkraść w przejście pomiędzy każdą z tych kwestii? Czy nie może być prawdą, że uważam jedną rzecz, ale z powodu błędnego przejścia między stanami *myślę* inną rzecz? (Jeśli możesz coś błędnie powiedzieć, to czy nie możesz czegoś błędnie pomyśleć?) Czy nie byłoby możliwe pomylić intencji albo wyrazić inną rzecz od tej, którą masz na myśli? A czy ułomna pamięć w podsystemie intencji komunikacyjnej nie może doprowadzić do tego, że rozpoczniesz wyrażanie pewnego komunikatu przedwerbalnego, a skończysz z innym komunikatem przedwerbalnym będącym standardem, wyznaczającym, co jest błędem do poprawienia? Między dwiema różnymi rzeczami istnieje logiczna przestrzeń na błąd, a gdy mnożymy indywidualne stany z określonymi treściami, odkrywamy – lub tworzymy – wiele źródeł błędów.

Jest bardzo kuszące, aby przeciąć ten węzeł gordyjski, deklarując, że *moja myśl (lub pogląd) o tym, jak to ze mną jest*, jest po prostu tym samym, czym *rzeczywiście jest moje przeżycie*. Innymi słowy, istnieje pokusa, by twierdzić, że logicznie nie ma żadnej przestrzeni na pojawienie się błędu między nimi, ponieważ są jedną i tą samą rzeczą. Takie twierdzenie ma pewne sympatyczne właściwości. Zatrzymuje zagrożenie eksplozją na kroku pierwszym – to zwykle świetne miejsce, aby powstrzymać eksplozję lub regres – i ma pewien naprawdę intuicyjny urok, świetnie widoczny w pytaniu retorycznym: Jaki sens mogłoby mieć twierdzenie, że coś jedynie wydawało mi się, że mi się wydawało (że mi się wydawało...), że było koniem?

Musimy jednak uważać, aby nie wdepnąć tu w pozostałości zapomnianych teorii filozoficznych (łącznie z niektórymi z moich – zob. Dennett 1969, 1978c, 1979a). Wydawałoby się, że możemy pozostać przy starych i dobrych kategoriach psychologii potocznej dotyczących przekonań, myśli, przekonań o przekonaniach, myśli o przeżyciach itp., i uniknąć rozdarcia wiedzy o sobie samym poprzez połączenie zwrotnych przypadków wyższego i niższego rzędu: deklarując, że gdy uważam, że uważam, że na przykład *p*, wynika z tego logicznie, że uważam, że *p*, i w tym samym duchu, gdy myślę, że mnie boli, to wynika z tego logicznie, że mnie boli itd.

Gdyby tak było, to na przykład gdy wyraziłem przekonanie drugiego rzędu, relacjonując przekonanie pierwszego rzędu, tak naprawdę po prostu miałbym do czynienia z jednym stanem, jedną rzeczą, a fakt, że relacjonując jedną rzecz, wyrażałem „inną", byłby złudzeniem powstałym na mocy li tylko rozróżnienia werbalnego, jak fakt, że Jones z początku chciał ożenić się ze swoją narzeczoną, ale w końcu ożenił się ze swoją żoną.

Jednak to połączenie nie wystarczy. Aby to dostrzec, po raz kolejny rozważmy rolę pamięci według psychologii potocznej. Nawet jeśli intuicyjnie jest przekonujące, że nie możesz się mylić co do tego, jak to jest z tobą *teraz*, zupełnie nie jest intuicyjnie przekonujące, że nie

możesz się mylić co do tego, jak z tobą było *wcześniej*. Jeśli przeżycie, które relacjonujesz, jest przeszłym doświadczeniem, twoja pamięć – na której opierasz się w swojej relacji – może zostać zanieczyszczona przez błąd. Być może twoje przeżycie wyglądało w jakiś sposób, ale teraz źle je pamiętasz jako takie, które wyglądało inaczej. Z pewnością może ci się wydawać *teraz*, że wydawało ci się *wówczas*, że to był koń – nawet jeśli *w rzeczywistości* wydawało ci się *wówczas*, że była to krowa. Logiczna możliwość błędnego pamiętania istnieje bez względu na to, jak krótka jest przerwa między rzeczywistym przeżyciem a następującym po nim przypomnieniem – to właśnie uprawniało teorie orwellowskie. Jednak w rozdziale 5 widzieliśmy, że błąd wkradający się do późniejszego przekonania dzięki orwellowskiemu fałszowaniu pamięci jest nieodróżnialny – zarówno z zewnątrz, jak i z wewnątrz – od błędu wkradającego się do pierwotnego przeżycia z powodu stalinowskiego pozoru. Więc nawet jeśli moglibyśmy podtrzymać to, że masz „bezpośredni" i „natychmiastowy" dostęp do twojej obecnej *oceny* (twojej myśli drugiego rzędu o tym, jakie rzeczy się tobie teraz wydają), tym samym nie będziesz w stanie wyeliminować możliwości, że jest to *błędna* ocena tego, jakie tobie się coś wydawało chwilę temu.

Jeśli indywidualizujemy stany (przekonania, stany świadomości, stany intencji komunikacyjnej itp.) ze względu na ich treść – co jest standardowym środkiem indywidualizacji w psychologii potocznej – będziemy musieli postulować różnice, które są systematycznie nie do odnalezienia żadnymi środkami, z wewnątrz bądź z zewnątrz, a przez to stracimy subiektywną bliskość czy niekorygowalność, która jest rzekomo charakterystyczną cechą świadomości. Widzieliśmy już tego przykłady w rozważaniach z rozdziału 5 dotyczących modeli zjawisk czasowych – orwellowskiego i stalinowskiego. A rozwiązaniem nie jest uczepienie się jednej bądź drugiej doktryny z psychologii potocznej, ale porzucenie tej jej cechy.

Zastępujemy podział na odrębne *stany* sensowne – przekonania, metaprzekonania itd. – *procesem* zapewniającym z czasem dobre dopasowanie wewnętrznych zdarzeń informacyjnych w jednostce do umiejętności jednostki *wyrażenia* (części) treści tych zdarzeń w mowie. Otto założył, że właśnie to miały zapewnić stany wyższego rzędu, ale one nie odzwierciedlały rzeczywistych podziałów w naturze. Linie podziału pojawiły się zupełnie nie tam, gdzie powinny.

Te wytwory psychologii potocznej nadal jednak żyją jako mieszkańcy światów heterofenomenologicznych podmiotów, których światopoglądy są rzeczywiście kształtowane przez ten schemat pojęciowy. Mówiąc tautologicznie, skoro ludziom rzeczywiście wydaje się, że posiadają zarówno te przekonania o swoich przeżyciach, jak i (dodatkowo) przeżycia same w sobie, te przeżycia i przekonania na ich temat należą do tego, jakie im się one wydają. Zatem musimy wyjaśnić *ten* fakt – nie fakt, że nasze umysły zawierają hierarchie reprezentacyjnych „stanów" przekonań wyższego rzędu, metaprzekonań itp., ale że *zwykle wydaje nam się*, że nasze umysły mają tego rodzaju organizację.

Przedstawiłem dwie hipotezy, dlaczego zwykle uważamy to za atrakcyjny pogląd. Po pierwsze, mamy nawyk zakładania istnienia oddzielnego procesu obserwacji (teraz *wewnętrznej* obserwacji) pośredniczącego między okolicznościami, które chcemy zrelacjonować, oraz samo sprawozdanie – a przeoczamy fakt, że w którymś momencie ten regres wewnętrznych obserwatorów musi zostać zatrzymany przez proces łączący treści z ich werbalnym wyrazem bez żadnego pośredniego oceniacza treści. Po drugie, wewnętrzna komunikacja stworzona w ten sposób rzeczywiście skutkuje nadaniem naszym umysłom organizacji w postaci nieskończenie potężnych refleksyjnych czy automonitorujących się systemów. Taką moc refleksji często nie bez powodu uważa się za kluczową cechę świadomości. Możemy wykorzystać uproszczony model psychologii potocznej jako swego rodzaju podporę wyobraźni, gdy próbujemy zrozumieć systemy automonitoringu, ale gdy go używamy, ryzykujemy powrót do kartezjańskiego

materializmu. Musimy zacząć uczyć się dawać sobie radę bez tej podpory i w następnym rozdziale zrobimy kilka dalszych ostrożnych kroków.

Rozdział 11

Demontowanie programu ochrony świadków

1. Podsumowanie

W części I zbadaliśmy problemy oraz wyłożyliśmy pewne metodologiczne założenia i zasady. W części II naszkicowaliśmy nowy model świadomości, model wielokrotnych szkiców, oraz zaczęliśmy uzasadniać, dlaczego powinnyśmy go woleć od modelu tradycyjnego, teatru kartezjańskiego. Podczas gdy idea teatru kartezjańskiego, zanalizowana bliżej, dość spektakularnie ujawnia swoje wady – nie istnieją zdeklarowani materialiści kartezjańscy – drugoplanowe założenia i nawyki myślowe, którym sprzyjał, nadal prowadzą do zarzutów i „fałszują" intuicję. Teraz w części III badamy następstwa naszego konkurencyjnego modelu, odpowiadając na serię przekonujących zarzutów. Niektóre z nich zdradzają nieustającą – nawet jeśli niejawną – wierność staremu, dobremu teatrowi kartezjańskiemu.

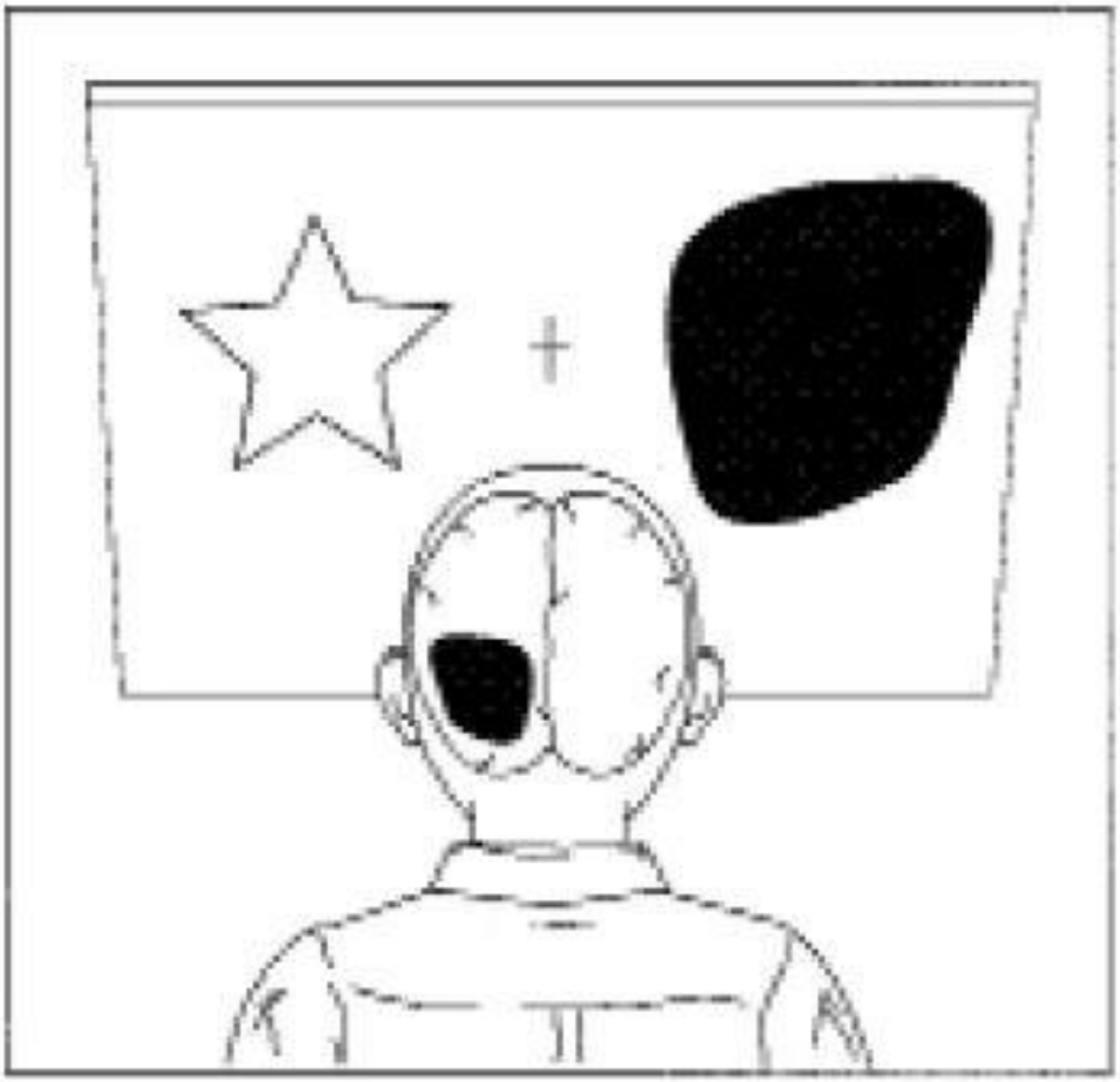
„Gdzie jednak następuje zrozumienie?" To pytanie stanowi istotę sporu od XVII wieku. Kartezjusz napotkał mur sceptycyzmu, gdy stwierdził (poprawnie), że mechanizmy w mózgu mogą wyjaśnić rozumienie przynajmniej w ogromnej mierze. Na przykład Antoine Arnauld w swoich obiekcjach do *Medytacji* uznał, że „wydaje się bowiem na pierwszy rzut oka nie do wiary, jak to się może dokonać bez udziału jakiejś duszy, aby światło odbite od ciała wilka poruszało najdelikatniejsze włókna nerwów wzrokowych owcy i dzięki temu poruszeniu, docierającemu aż do mózgu, rozchodziło się po nerwach zmysłowych tchnienie żywotne w sposób taki, jaki jest konieczny do tego, by owca porwała się do ucieczki" (Descartes 1641/2001, s. 182). Kartezjusz odpowiedział, że nie jest to bardziej nie do wiary od ludzkiej umiejętności wyrzucenia rąk przed siebie w ochronie przed upadkiem, co również jest reakcją mechaniczną, bez udziału „duszy". Ta idea „mechanicznej" interpretacji w mózgu jest kluczowym spostrzeżeniem *każdej* materialistycznej teorii umysłu, ale podważa głęboko tkwiące przeczucie: nasze poczucie, że aby nastąpiło *prawdziwe* zrozumienie, musi być *tam ktoś*, kto uprawomocni jego przebieg, kto *będzie świadkiem* rzeczy, których zachodzenie stanowi o zrozumieniu. (Filozof John Searle wykorzystuje to poczucie w swoim słynnym eksperymencie myślowym Chiński Pokój, którym zajmiemy się w rozdziale 14).

Kartezjusz był mechanicystą w całym znaczeniu tego słowa, w kwestii każdego zjawiska w naturze, ale w przypadku ludzkiego umysłu miał opory. Twierdził, że oprócz mechanicznej interpretacji mózg stanowi też budulec dla centralnej sfery – którą ja nazywam „teatrem kartezjańskim" – gdzie, w istotach ludzkich, dusza może być świadkiem i dokonywać własnych osądów. Świadkowie potrzebują surowców, na których mogą oprzeć osądy. Te surowce, bez względu na to, czy nazywane są „danymi zmysłowymi", „wrażeniami", „surowymi czuciami" czy „fenomenalnymi własnościami przeżyć", są rekwizytami, bez których istnienie świadka nie ma sensu. Te rekwizyty, pozostające na mocy różnych złudzeń, otaczają ideę centralnego świadka barierą intuicji niemal nie do pokonania. Zadaniem tego rozdziału jest przebić się przez tę barierę.

2. Ślepowidzenie: częściowa zombifikacja?

Tylko znikomy odsetek wszystkich strasznych wypadków przydarzających się ludziom zostaje częściowo odkupiony, gdyż odkrywa przed dociekliwymi naukowcami sekrety natury. Dotyczy to w szczególności uszkodzeń mózgu spowodowanych traumą (strzelaniną, wypadkami drogowymi itp.), nowotworem czy udarem[101]. Powstałe w ich skutek wzorce upośledzenia i zachowanych umiejętności czasem stanowią istotne – a nawet zdumiewające – świadectwo tego, jak umysł jest realizowany przez mózg. Jednym z najbardziej zaskakujących, jak wskazuje jego paradoksalna nazwa, jest ślepowidzenie. Z początku wydaje się, że zostało ono stworzone na zamówienie eksperymentów myślowych filozofów: przypadłość zamieniająca normalną, świadomą osobę w częściowego zombi, nieświadomy automat w odniesieniu do pewnych bodźców, ale osobę zupełnie świadomą w innych kwestiach. Nie jest więc zaskoczeniem, że filozofowie nadali ślepowidzeniu rodzaj mitycznego statusu, jako przykład, wokół którego można snuć kolejne przykłady. Jak jednak zobaczymy, ślepowidzenie nie wspiera pojęcia zombi; podważa go.

W przypadku normalnego ludzkiego widzenia przychodzące sygnały z oczu podróżują nerwami wzrokowymi przez różne stacje pośrednie do płatu potylicznego czy kory wzrokowej, czyli do części mózgu w samym tyle czaszki, nad móżdżkiem. Informacje o lewym polu widzenia (o lewych połowach pól z każdego oka) zostają rozprzestrzenione w prawej połowie kory wzrokowej, a o prawym polu widzenia – w lewej połowie. Od czasu do czasu wypadek naczyniowy (np. pęknięcie naczynia krwionośnego) niszczy część płatu potylicznego, tworząc plamkę ślepą, czyli ubytek pola widzenia (mroczek), stosunkowo sporą dziurę w świecie przeżywanym wzrokowo, po stronie naprzeciw uszkodzenia.



Ryc. 11.1

W skrajnych przypadkach, gdzie zarówno lewa, jak i prawa część kory wzrokowej zostały zniszczone, osoba zostaje zupełnie niewidoma. Częściej cała kora wzrokowa po jednej stronie mózgu zostaje zniszczona przez wypadek naczyniowy, prowadząc do utraty przeciwnej połowy pola widzenia; utrata lewej połowy kory wzrokowej powoduje niedowidzenie połowiczne prawostronne, całkowitą ślepotę w prawej połowie pola widzenia.

Jak to jest mieć mroczek w polu wzrokowym? Mogłoby się wydawać, że wszyscy wiemy, gdyż wszyscy mamy ślepe plamki w polach widzenia, odpowiadające miejscom na naszej siatkówce, gdzie nie ma pręcików ani czopków, ponieważ nerw wzrokowy wychodzi w tym miejscu z gałki ocznej. Zwykła ślepa plamka, czy też tarcza nerwu wzrokowego, nie jest mała: wymazuje koło o średnicy około 6 stopni pola widzenia. Zamknij jedno oko i popatrz na krzyżyk, trzymając stronę w odległości około 25 centymetrów od oczu. Jedno z kół „ślepej plamki" powinno zniknąć. Zamknij drugie oko, a wówczas powinno zniknąć drugie koło. (Może będziesz

musiał dopasować odległość od kartki, aby dostrzec ten efekt. Cały czas patrz prosto na krzyżyk). Dlaczego zwykle nie dostrzegasz tej luki w polu widzenia? Częściowo dlatego, że masz dwoje oczu i jedno z nich kryje drugie; ich ślepe plamki nie zachodzą na siebie. Jednak nawet z jednym okiem zamkniętym w większości warunków nie zauważysz swojej ślepej plamki. Dlaczego nie? Twój mózg nigdy nie musiał radzić sobie z informacją wejściową z tego obszaru siatkówki, więc nie poświęcił żadnych środków na to, aby się tym zająć. Nie ma homunkulusów odpowiedzialnych za otrzymywanie raportów z tego obszaru, więc gdy nie docierają żadne raporty, nikt nie narzeka. Brak informacji nie jest tym samym co informacja o braku. Aby zobaczyć lukę, coś w twoim mózgu musiałoby odpowiedzieć na kontrast: *albo* pomiędzy brzegiem wewnętrznym i zewnętrznym – a twój mózg nie ma maszynerii, aby robić to w tym miejscu – *albo* pomiędzy przed i po: teraz widzisz koło, a teraz nie. (W taki sposób znikające, czarne koło na rycinie 11.2 informuje cię o ślepej plamce).

Ryc. 11.2

Tak jak nasze zwykłe plamki ślepe, luki w polach wzrokowych mają konkretne lokalizacje, a niektóre – ostre granice, które mogą być łatwo oznaczone przez eksperymentatora przy użyciu bodźca takiego jak punkt świetlny poruszany w polu widzenia badanego. Badany zostaje poproszony o zrelacjonowanie, kiedy punkt nie jest już przeżywany – odmiana eksperymentu, który właśnie przeprowadziliśmy na sobie w celu odkrycia plamki ślepej. Relacje badanych mogą następnie zostać skorelowane z mapami zniszczeń w korze, wytworzonymi przez skany mózgu z CT (tomografia komputerowa) oraz MRI (obrazowanie metodą rezonansu magnetycznego). Mroczek różni się pod pewnym istotnym względem od zwykłej plamki ślepej: zwykle jest dostrzegany przez badanego. Nie wynika to tylko z tego, że jest większy niż zwykła plamka ślepa. Jest spowodowany utratą komórek w korze wzrokowej, które wcześniej „informowały" inne komórki w korze, które również „troszczyły się" o informacje z pewnych obszarów siatkówki, więc ich nieobecność zostaje zauważona. Oczekiwania mózgu są zdezorganizowane; brakuje czegoś, co się tam powinno znajdować, jakiś głód epistemologiczny pozostaje niezaspokojony. Zatem osoba badana jest zwykle świadoma luki w polu wzrokowym, ale jako braku, a nie jako pozytywnie czarnego regionu, który można by dostrzec, gdyby ktoś przykleił okrągły kawałek czarnego papieru na przedniej szybie twojego samochodu.

Skoro normalne ścieżki wzrokowe w mózgu zostały zakłócone lub ucięte, można by się spodziewać, że osoby posiadające luki w polu wzrokowym zupełnie nie będą w stanie zebrać

informacji na temat rzeczy zachodzących w ich ślepych polach. W końcu są one ślepe. A mówiąc dokładnie: nie przeżywają nic wizualnego ani wewnątrz, ani na obrzeżach obszaru ślepego – żadnych błysków, krawędzi, barw, migotania czy gwiazdek. Niczego. Tym właśnie jest ślepota. Jednak niektórzy ludzie z tym schorzeniem przejawiają niezwykły talent: mimo zupełnego braku świadomego przeżycia wizualnego w ślepym obszarze mogą czasem „odgadnąć" z niesamowitą dokładnością, czy właśnie zaświeciło się światło w ślepym polu, czy nie, a nawet, czy pokazany został kwadrat, czy koło. Jest to zjawisko zwane ślepowidzeniem (Weiskrantz 1986, 1988, 1990). Sporne pozostaje wyjaśnienie ślepowidzenia, ale żaden badacz nie uważa, że dzieje się tam coś „paranormalnego". Istnieje przynajmniej dziesięć różnych ścieżek pomiędzy siatkówką i resztą mózgu, więc nawet jeśli zniszczony jest płat potyliczny, nadal pozostaje mnóstwo kanałów komunikacyjnych, którymi informacje z zupełnie normalnych siatkówek mogą dotrzeć do innych obszarów mózgu. Przeprowadzono wiele testów na pacjentach ze ślepowidzeniem i nie ma wątpliwości, że częściej niż przypadkiem (a nawet w stu procentach przypadków w pewnych warunkach) są oni w stanie odgadnąć różne proste kształty, kierunek ruchu, obecność lub brak światła. Żadna osoba cierpiąca na ślepowidzenie nie wykazała się jeszcze umiejętnością rozróżniania barw w ślepym polu, lecz niedawne badania przeprowadzone przez Stoerig i Coweya (1990) dostarczają świadectw, że to jest możliwe.

Jak działa ślepowidzenie? Czy jest to, jak twierdzą niektórzy filozofowie i psychologowie, percepcja wizualna bez świadomości – coś w rodzaju tego, co może mieć zwykły automat? Czy jest ona świadectwem (a przynajmniej poważnym kłopotem) przemawiającym przeciwko funkcjonalistycznym teoriom mózgu, gdyż ukazuje przypadek, w którym wszystkie *funkcje* wzroku są nadal obecne, ale zostały odcedzone z całego smakowitego soczku *świadomości*? Niczego takiego nie dowodzi. W pośpiechu, aby zaprząc ślepowidzenie do swoich ideologicznych wozów, filozofowie momentami przeoczyli pewne raczej elementarne fakty dotyczące zjawiska ślepowidzenia oraz eksperymentalnego otoczenia, w którym się ujawnia.

Tak jak w przypadku „czasowych anomalii" analizowanych w rozdziałach 5 i 6, zjawisko ślepowidzenia pojawia się jedynie, gdy traktujemy badanych z punktu widzenia heterofenomenologii. Eksperymenty nie mogłyby zostać przeprowadzone, gdyby eksperymentatorzy nie mogli dać werbalnych instrukcji badanym (i upewnić się, że zostali zrozumiani), a odpowiedzi badanych dostarczają świadectw na rzecz tego zaskakującego zjawiska *tylko* wówczas, gdy są one interpretowane jako akty mowy. Jest to coś niemalże zbyt oczywistego, aby dało się dostrzec, więc muszę zrobić przerwę na wyjaśnienia.

Interpretacja ślepowidzenia jest w wielu aspektach sporna, lecz zupełnie bezdyskusyjna pod jednym względem: wszyscy zgadzają się, że pacjent ze ślepowidzeniem jakoś dowiaduje się o pewnych wydarzeniach ze świata przez oczy (to część „widzenie"), mimo iż nie ma świadomego przeżycia wzrokowego odpowiedniego zdarzenia (to część „ślepo"). Mówiąc prościej, ślepowidzenie zakłada (1) przyjęcie informacji wizualnych, które są (2) mimo wszystko nieświadome. Dowód na (1) jest banalny: badany częściej niż przypadkowo daje sobie radę na testach dotyczących takich informacji. Dowód na (2) jest bardziej poszlakowy: badani zaprzeczają, jakoby byli świadomi takich wydarzeń, a ich werbalne zaprzeczenia mają na swoje poparcie świadectwa neurologiczne związane z uszkodzeniem mózgu z jednej strony oraz spójność ich zaprzeczeń z drugiej strony. Więc im wierzymy![102]

Nie jest to kwestia banalna. Zauważmy, że to, co jest uderzające w ślepowidzeniu, znikłoby natychmiast, gdybyśmy stwierdzili, iż badani symulują ten stan – jedynie udają, że nie są świadomi. Albo, co interesuje nas bardziej, porównajmy akceptację zaprzeczeń pacjentów ze ślepowidzeniem do sceptycyzmu, z którym podchodzimy do takich samych zaprzeczeń

pochodzących od ludzi ze stwierdzoną „ślepotą histeryczną". Czasem ludzie, których oczy i mózgi wydają się pracować poprawnie, o ile potrafią to stwierdzić psychologowie, mimo wszystko twierdzą, że oślepli; uzasadniają to twierdzenie, zachowując się „jak osoba niewidoma". Można by znaleźć dosyć wiarygodny powód, który skłania te osoby do „stania się" niewidomymi – to kara dla nich samych lub dla kogoś, kto teraz musi się martwić i im współczuć, bądź zaprzeczenie pewnemu okropnemu wspomnieniu wizualnemu, czy też rodzaj paniki w odpowiedzi na jakąś inną chorobę czy osłabienie – więc jest to ślepota „psychosomatyczna", jeśli w ogóle jest ślepotą. Czy ci ludzie rzeczywiście są niewidomi? Mogą być. W końcu można by powiedzieć, że jeśli ból psychosomatyczny jest prawdziwym bólem, a psychosomatyczne nudności są wystarczająco prawdziwe, aby zwymiotować, to dlaczego ślepota psychosomatyczna nie miałaby być prawdziwą ślepotą?

Ludzie ze ślepotą histeryczną *twierdzą*, że są ślepi, ale jak ci ze ślepowidzeniem, mimo wszystko dają niezbite dowody na to, że *przyjmują informacje wizualne*. Na przykład osoby ze ślepotą histeryczną zwykle zdecydowanie *rzadziej* niż przypadkiem radzą sobie z prośbą o odgadnięcie wizualnych własności rzeczy! Jest to pewny znak, że w jakiś sposób używają informacji wizualnych, aby w zachowaniu przeważały „błędy". Ludzie ci mają niespotykany talent odnajdywania krzeseł, na które mogą wpaść. Mimo to, w przeciwieństwie do oczywistego symulowania, gdy mówią, że nie mają przeżyć wizualnych, są szczerzy – naprawdę w to wierzą. Czy nie powinniśmy wierzyć i my? Jak powinniśmy traktować teksty tych dwóch różnych grup badanych, gdy wnioskujemy o ich światach heterofenomenologicznych?

Oto miejsce, w którym niezwykle ostrożna polityka heterofenomenologii procentuje. Zarówno osoby ze ślepowidzeniem, jak i ze ślepotą histeryczną są najwyraźniej szczere w swoich deklaracjach, że są nieświadome czegokolwiek, co dzieje się w ich ślepych polach. Zatem ich światy heterofenomenologiczne są podobne – przynajmniej w kwestii domniemanych ślepych pól. A jednak jest między nimi różnica. Mamy mniejszą wiedzę o neuroanatomicznej podbudowie ślepoty histerycznej niż ślepowidzenia, a jednak intuicyjnie czujemy się bardziej sceptyczni w stosunku do ich zaprzeczeń[103]. Co wzbudza podejrzenie, że histerycznie ślepi nie są *naprawdę* ślepi, a nawet że są w pewnym sensie lub do pewnego stopnia świadomi swoich światów wizualnych. Podejrzanie pomyślne warunki ich ślepoty sprawiają, że się dziwimy, ale poza tym dowodem poszlakowym jest prostszy powód: wątpimy w ich twierdzenia o ślepocie, ponieważ *bez zachęty* ludzie ze ślepotą histeryczną czasem *korzystają* z informacji pochodzących z oczu w sposób, w jaki nie robią tego ludzie ze ślepowidzeniem.

W sytuacjach eksperymentalnych ze ślepowidzeniem obecny jest czynnik, który tak świetnie wpasowuje się w nasze standardowe założenia, że niemal nikt nie zajmuje się jego omówieniem (ale zob. Marcel 1988; van Gulick 1989; Carruthers 1989): badani ze ślepowidzeniem muszą zostać zachęceni, aby częściej niż przypadkiem „odgadywali". Dlatego eksperymentator może powiedzieć w początkowej instrukcji: „gdy tylko usłyszysz sygnał, zgaduj" lub „gdy tylko poczujesz, że klepię cię w rękę, daj odpowiedź". Bez takiej zachęty badani po prostu nie reagują[104].

Możemy przetestować naszą diagnozę tej różnicy wyobrażając sobie pewną wariację. Załóżmy, że spotkaliśmy osobę z rzekomym ślepowidzeniem, która nie potrzebowała sygnału: „spontanicznie" przekazuje swoje „odgadnięcia" (zdecydowanie częstsze niż przypadkowe, ale nie idealne) tylekroć, ilekroć coś zostaje zaprezentowane w rzekomym polu ślepym. Sadzamy ją w laboratorium i przeprowadzamy zwykły test, aby stworzyć mapę domniemanej luki w polu wzrokowym; mówi nam, kiedy poruszające się światło znika w jej ślepym polu, tak jak każdy inny pacjent ze ślepowidzeniem. Jednocześnie jednak spontanicznie, bez żadnej zachęty, mówi na przykład: „Tylko zgaduję, ale czy właśnie zaświeciłeś światło w moim obszarze ślepym?" –

ale jedynie wówczas, gdy właśnie to zrobiliśmy. Byłoby to, delikatnie mówiąc, podejrzane i możemy powiedzieć dlaczego.

Ogólnie rzecz biorąc, gdy badani wykonują instrukcje podczas eksperymentu, jest to uznawane za jasny dowód na to, że są w stanie je wykonywać, ponieważ *świadomie przeżyli* odpowiednie zdarzenia związane z bodźcem. Dlatego poniższa instrukcja przygotowująca zostałaby uznana za bezsensowną:

Gdy będziesz mieć *świadomość* zaistnienia światła, wciśnij lewy przycisk; gdy światło będzie się nadal świeciło, ale *nie* będziesz mieć tej świadomości, wciśnij prawy przycisk.

Jak, u licha, badany miałby wypełnić tę instrukcję? Prosilibyśmy badanego o coś niewykonalnego: uzależnienie swojego zachowania od zdarzeń dla niego niedostępnych. To tak, jakby powiedzieć: „podnieś rękę, gdy ktoś puści do ciebie oko bez twojej wiedzy". Badacz nie czułby potrzeby dodania przysłówka „świadomie", jak w przypadku: „Gdy świadomie usłyszysz sygnał, zgaduj", gdyż standardowe założenie jest takie, że nie można uzależniać swojego działania od nieświadomych przeżyć, nawet jeśli takie rzeczy się zdarzają. Aby przyjąć politykę „Gdy wydarzy się *x*, zrób *y*", musisz móc być świadomym, że dzieje się *x*.

Jest to nasze typowe założenie, jednak ten gmach oczywistości ma pewne pęknięcie. Czyż nie wiemy, że wiele naszych zachowań zależy od warunków, które wykrywamy jedynie nieświadomie? Pomyślmy o strategii regulującej temperaturę ciała, przystosowującej metabolizm, przechowującej i odzyskującej energię, aktywującej układy immunologiczne; pomyślmy o strategiach takich jak mruganie, gdy obiekty zbliżają się do oka lub do niego wpadają, a nawet publiczne zachowania działające na tak dużą skalę, jak chodzenie (bez przewracania się) i uchylanie się, gdy jakiś obiekt nagle się do nas zbliża. Całe to „zachowanie" jest kontrolowane bez żadnego udziału świadomości – jak zauważył Kartezjusz.

Wydaje się zatem, że istnieją dwa rodzaje strategii behawioralnych: kontrolowane przez świadomą myśl oraz kontrolowane przez „ślepe, mechaniczne" procesy – jak procesy kontrolujące zautomatyzowaną windę. Jeśli winda musi trzymać się strategii przewożenia nie więcej niż tony, musi mieć jakiś rodzaj wbudowanej wagi, aby wykryć, kiedy ten limit zostaje przekroczony. Winda z pewnością nie jest świadoma ani niczego nie wykrywa świadomie, a zatem nie ma świadomych strategii. Można jednak powiedzieć, że stosuje strategie dotyczące różnych stanów w świecie, które to stany wykrywa, a nawet że dostosowuje stosowane strategie na podstawie innych, wykrywanych przez nią stanów rzeczy itd. Może posiadać strategie, metastrategie oraz meta-metastrategie, wszystkie dotyczące różnych skomplikowanych kombinacji wykrytego stanu rzeczy – i wszystko to bez cienia świadomości. Cokolwiek może zrobić winda, aby wykrywać i stosować strategie, z pewnością również mogą dokonać ludzki mózg i ciało. Mogą stosować rozbudowane, nieświadome strategie typu windowego.

Jaka jest więc różnica między nieświadomym stosowaniem strategii a świadomym stosowaniem strategii? Gdy rozważymy strategie, których nasze ciała używają nieświadomie dzięki „ślepym, mechanicznym" detektorom warunków, kuszące jest stwierdzenie, że *skoro* są to strategie nieświadome, są nie tyle *naszymi* strategiami, ile strategiami *naszych ciał*. *Nasze* strategie są (można by powiedzieć: z definicji) naszymi świadomymi strategiami; tymi, które *my* świadomie i celowo tworzymy, mając możliwość (świadomego) zastanowienia się nad ich słabymi i mocnymi stronami oraz dostosowania czy poprawienia ich, gdy sytuacja w naszym przeżyciu będzie tego wymagać.

Wydaje się zatem, że gdy jakaś strategia jest początkowo przyjęta w wyniku werbalnego ustalenia lub w odpowiedzi na werbalną instrukcję, jest tym samym strategią świadomą, z konieczności dotyczącą świadomie przeżytych zdarzeń (Marcel 1988). Wewnętrznie sprzeczne jednak wydaje się to, że można by o tym porozmawiać i wówczas zdecydować się na

wypełnianie *nie*świadomej strategii, dotyczącej nieświadomie wykrytych zdarzeń. Istnieje jednak furtka: status takiej strategii *może* się zmienić. Z wystarczającą praktyką i odrobiną strategicznie rozłożonego zapominalstwa możemy zacząć od świadomie przyjętej i wypełnianej strategii i stopniowo przesuwać się do stanu wypełniania nieświadomej strategii, wykrywając odpowiednie elementy bez ich świadomości. Mogłoby się to wydarzyć, ale tylko wówczas, gdyby połączenie z werbalnym przemyśleniem strategii zostało w jakiś sposób zerwane.

To możliwe przejście można lepiej ukazać w przeciwnym kierunku. Czy pacjent ze ślepowidzeniem nie mógłby uświadomić sobie wizualnych przeżyć w obszarze ślepym przez odwrócenie przed chwilą wyobrażonych procesów? W ślepowidzeniu mózg badanego wyraźnie przecież otrzymuje i analizuje informacje wizualne, które w jakiś sposób są wykorzystywane do dobrego zgadywania. Krótko po tym, jak pojawi się impuls w mózgu pacjenta, następuje coś, co daje początek stanowi *poinformowania*. Jeśli zewnętrzny obserwator (taki jak eksperymentator) może doprowadzić do rozpoznania tego początku, mógłby on zasadniczo przekazać informację badanemu. W ten sposób badany mógłby rozpoznać te początki „z drugiej ręki", mimo że nie byłby ich świadom „bezpośrednio". A potem czy osoba badana nie powinna móc w zasadzie „wyeliminować pośrednika" i rozpoznać, tak jak to robi eksperymentator, zmiany w swoich własnych dyspozycjach? Z początku mogłoby to wymagać użycia jakiegoś rodzaju sprzętu automonitorującego – tego samego, z którego korzysta eksperymentator – ale badany mógłby teraz obserwować sygnały wyjściowe lub ich słuchać[105].

Innymi słowy, czy nie powinno w zasadzie być możliwe „zamknięcie pętli informacji zwrotnej", a w ten sposób wyszkolenie badanego, aby stosował strategię uzależniania swojego zachowania od zmian, których nie przeżył („bezpośrednio")? Mówię o perspektywie takiego szkolenia dla ślepowidzących tak, jakby był to jedynie eksperyment myślowy, ale tak naprawdę mógłby on zostać bardzo łatwo zmieniony w prawdziwy eksperyment. Moglibyśmy spróbować wyszkolić badanego ze ślepowidzeniem tak, aby wiedział, kiedy ma „zgadywać".

Talenty i dyspozycje pacjentów ze ślepowidzeniem nie są niezmienne; w niektóre dni są oni w lepszej formie niż w inne; ich wyniki są tym lepsze, im więcej ćwiczą, mimo że zwykle *nie* są natychmiast informowani przez eksperymentatora, jak dobrze sobie radzą (wyjątki zobacz w Zihl 1980, 1981). Istnieje ku temu kilka powodów, a główny z nich to taki, że w każdej takiej sytuacji eksperymentalnej występuje możliwość niezamierzonych i niezauważonych podpowiedzi ze strony eksperymentatora, więc interakcje między nim a badanym są skrupulatnie minimalizowane i kontrolowane. Mimo to badani karmią się sugestiami otrzymywanymi od eksperymentatora i stopniowo przyzwyczajają się do w innym razie dziwnie beznadziejnej praktyki zgadywania po sto i tysiąc razy rzeczy, co do których są przekonani, że nie mają żadnych przeżyć. (Wyobraźmy sobie, jak byśmy się czuli, gdyby poproszono nas, abyśmy usiedli z książką telefoniczną i zgadli, jakiej marki samochód posiada każda ze znajdujących się tam osób, a nikt nie mówiłby nam, czy strzały są poprawne. Nie wydawałoby się to zbyt długo celowe, chyba że dostalibyśmy wiarygodne zapewnienie na temat tego, jak nam idzie oraz dlaczego jest to wyczyn, którym warto się zajmować).

Cóż by się zatem stało, gdybyśmy odsunęli inne cele naukowe i zobaczyli, ile moglibyśmy osiągnąć, szkoląc kogoś ze ślepowidzeniem, korzystając z takiego rodzaju informacji zwrotnej, która wydawałaby się pomocna? Załóżmy, że zaczynamy ze standardowym pacjentem cierpiącym na ślepowidzenie, który „odgaduje", gdy go o to prosimy (tzw. metoda wymuszonego wyboru), i którego poprawne odpowiedzi są częstsze niż przypadkowe (jeśli nie są, to nie jest osobą ze ślepowidzeniem). Informacja zwrotna szybko sprawiłaby, że odpowiedzi prawidłowe stałyby się możliwie najczęstsze, a gdyby odgadywanie ustabilizowało się na pewnym wysokim poziomie poprawności, powinno to zaimponować badanemu, który

stwierdziłby, że ma użyteczny i niezawodny talent i być może warto go rozwijać. Jest to faktycznie stan, w którym niektórzy badani są dziś.

Teraz załóżmy, że zaczynamy prosić badanego, aby działał bez podpowiedzi – aby „zgadywał, kiedy zgadywać", aby zgadywał, „gdy tylko będzie miał na to ochotę" – i ponownie załóżmy, że eksperymentator dostarcza natychmiastową informację zwrotną. Są dwa możliwe wyniki:

(1) Badany zaczyna zgadywać przypadkowo i nie poprawia się. Mimo że jest on wyraźnie informowany przez początek zdarzeń bodźcowych, wydaje się, iż nie sposób, by mógł on odkryć, kiedy nastąpiło owo poinformowanie, bez względu na to, jakiego podparcia „biologicznego sprzężenia zwrotnego" mu dostarczymy.

(2) Badany w końcu jest w stanie pracować bez podpowiedzi eksperymentatora (ani żadnego tymczasowego wsparcia biologicznego sprzężenia zwrotnego) i utrzymuje się na poziomie znacząco wyższym niż przypadkowy.

Który z rezultatów otrzymalibyśmy w każdym przypadku, to oczywiście kwestia empiryczna i nie zamierzam nawet próbować odgadnąć, jak prawdopodobny mógłby być rezultat drugiego typu. Być może w każdym przypadku badany nie byłby w stanie nauczyć się prawidłowo „odgadywać", kiedy odgadywać. Jednak zwróćmy uwagę, że gdyby miał wystąpić wynik drugiego typu, badany mógłby wówczas dosyć racjonalnie zostać poproszony o przyjęcie strategii zachowania względem bodźców, których pojawienie się mógł jedynie odgadnąć. Bez względu na to, czy był świadom tych bodźców, jeśli niezawodność jego „odgadywania" była wysoka, mógł traktować owe bodźce tak samo jak jakiekolwiek świadome przeżycie. Mógł rozmyślać i decydować o strategiach dotyczących ich występowania z taką samą łatwością jak w przypadku zdarzeń przeżywanych świadomie.

Ale czy w jakiś sposób *sprawiłoby* to, że uświadomiłby sobie bodźce? Co podpowiada ci intuicja? Gdy pytam ludzi, ku czemu skłanialiby się w takim przypadku, dostaję różne odpowiedzi. Psychologia potoczna nie daje jasnego werdyktu. Jednak pewien pacjent ze ślepowidzeniem opowiedział o podobnej sytuacji. D.B., jeden z pacjentów badanych przez Weiskrantza, cierpi na niedowidzenie połowiczne prawostronne i wykazuje umiejętność typową dla ślepowidzenia, czyli otrzymując podpowiedź, zgaduje częściej niż przypadkiem. Jeśli na przykład światło jest powoli poruszane przez jego obszar ślepy poziomo lub pionowo i otrzymuje on podpowiedź, aby zgadywał „pionowe czy poziome", fantastycznie daje sobie radę, jednocześnie zaprzeczając jakiejkolwiek świadomości ruchu. Lecz jeśli światło porusza się szybciej, samo w sobie staje się podpowiedzią: D.B. może bez zachęty zdać bardzo precyzyjną relację z ruchu, a nawet pokazać ten ruch swoją ręką, gdy tylko się pojawi (Weiskrantz 1988, 1989). A gdy zostanie o to zapytany, D.B. twierdzi, że *oczywiście* świadomie przeżył ruch – w jaki inny sposób byłby w stanie o nim opowiedzieć? (Inni badani ze ślepowidzeniem również mówią o świadomym przeżywaniu szybko poruszających się bodźców). Powinniśmy powstrzymać się od osądzania, jednak jego odpowiedź nie powinna nas dziwić, jeśli analiza codziennego pojęcia świadomości Rosenthala jest na dobrym tropie. D.B. nie zostaje po prostu poinformowany o ruchu światła; *zdaje sobie sprawę* z tego, że został poinformowany; używając terminów Rosenthala, ma myśl drugiego rzędu dotyczącą tego, że właśnie miał myśl pierwszego rzędu.

Powraca Otto, nasz krytyk:

Ale to tylko kolejny trik! Od zawsze wiemy, że pacjenci ze ślepowidzeniem są świadomi swojego zgadywania. Pokazuje to jedynie, że badany może rozwinąć talent do odgadywania,

kiedy odgadywać (i oczywiście byłby świadomy *tego* odgadywania). Rozpoznanie, że czyjeś odgadywanie na te tematy jest niezawodne, byłoby samo w sobie ledwo wystarczalne, aby ktoś mógł bezpośrednio uświadomić sobie zdarzenia, *które* odgaduje.

Potrzebujemy zatem czegoś więcej dla świadomości wizualnej. Co można dodać? Przede wszystkim połączenie między odgadnięciem i stanem, którego *dotyczy*, nawet jeśli jest niezawodne, to wydaje się dosyć słabe i efemeryczne. Czy może zostać poszerzone i wzmocnione? Jaki byłby rezultat, gdyby więzy dotyczenia między zgadnięciem a jego przedmiotem zostały zwielokrotnione?

3. Ukryć naparstek: ćwiczenie w uświadamianiu

Standardowe pojęcie filozoficzne na oznaczenie dotyczenia czegoś to *intencjonalność*, a według Elizabeth Anscombe (1965) „przychodzi ono przez metaforę" z łaciny, *intendere arcum in*, co oznacza *wycelować łuk i strzałę w* (coś). Ten obraz celowania i skierowania jest kluczowy w większości dysput filozoficznych o intencjonalności, jednak ogólnie rzecz biorąc, filozofowie zamienili skomplikowany proces celowania prawdziwą strzałą w zwykłą strzałę „logiczną", podstawową czy pierwotną relację, która jest jeszcze bardziej tajemnicza przez swoją domniemaną prostotę. Jak coś w twojej głowie *mogłoby* wycelować tę abstrakcyjną strzałę w coś w świecie?[106] Pojmowanie relacji dotyczenia jako abstrakcyjnej, logicznej relacji może koniec końców być poprawne, lecz z początku odwraca uwagę od procesów rzeczywiście związanych z utrzymywaniem umysłu w wystarczającym kontakcie z rzeczami ze świata, tak aby można o nich było myśleć *skutecznie*: procesów kierowania uwagi, bycia w kontakcie, śledzenia i tropienia (Selfridge, nieopublikowane). Rzeczywista kwestia celowania w coś, „trzymania czegoś na celowniku", zakłada przeprowadzenie w czasie serii regulacji i kompensacji pod „kontrolą informacji zwrotnej". Dlatego właśnie obecność czynników dekoncentrujących (takich jak folie zakłóceniowe, które mylą systemy antyrakietowe) może sprawić, że celowanie będzie niemożliwe. Utrzymanie celu na tyle długo, że można go zidentyfikować, jest osiągnięciem wymagającym czegoś więcej niż jednorazowej, chwilowej operacji informacyjnej. Najlepszym sposobem podtrzymywania z czymś kontaktu jest, dosłownie, *dotykanie* tego – chwycenie i niedopuszczenie do tego, aby uciekło, żeby badać to tak długo, jak mamy na to ochotę. Kolejnym świetnym sposobem utrzymania z czymś kontaktu przenośnego jest śledzenie tego oczami (i resztą ciała) bez spuszczania z tego wzroku. Można to zrobić percepcyjnie, ale nie tylko przez percepcję bierną; pozostanie z czymś w kontakcie może wymagać pewnego wysiłku, planowania i na pewno *bezustannej aktywności*.

Gdy byłem dzieckiem, uwielbiałem bawić się w dziecięcą grę *Ukryć naparstek*. Zwykły naparstek zostaje pokazany uczestnikom i wszyscy poza jednym opuszczają pomieszczenie, po czym naparstek zostaje „ukryty". Zasady dla osoby chowającej są jasne: naparstek musi zostać ukryty *na widoku*. Nie może zostać umieszczony za czymś lub pod czymś, lub też za wysoko, aby dzieci go zobaczyły. W przeciętym salonie są dziesiątki miejsc, w których można umieścić naparstek, gdzie zwykle zleje się on z otoczeniem jak dobrze zakamuflowane zwierzę. Gdy już jest schowany, reszta dzieci wraca do pokoju i rozpoczyna polowanie na naparstek. Gdy któreś tylko go dostrzeże, cicho siada, starając się nie zdradzić jego umiejscowienia. Kilkoro dzieci, które znajdują go jako ostatnie, może zwykle być pewnych, że kilka razy *spojrzało nań dokładnie* bez *dostrzeżenia* go. W tych wspaniałych momentach wszyscy widzą, że naparstek jest wprost przed nosem na przykład Betsy, świetnie i pod idealnym kątem widoczny w jej polu widzenia. (W takich momentach moja mama lubiła mówić: „Gdyby był niedźwiadkiem, toby cię ugryzł!") Przez chichot i westchnienia innych dzieci Betsy w końcu zdaje sobie sprawę, że musi patrzeć

wprost na niego – i nadal go nie widzi.

Można to ująć tak: nawet jeśli pewien reprezentacyjny stan w mózgu Betsy w jakiś sposób „obejmuje" naparstek, żaden jej stan percepcyjny jeszcze go nie *dotyczy*. Możemy przyznać, że jeden z jej świadomych stanów dotyczy naparstka: jej „obraz poszukiwany". Być może usilnie koncentruje się na znalezieniu go, czyli tego samego naparstka, który oglądała minutę czy dwie temu. Jednak żadna silna relacja intencjonalności czy dotyczenia czegoś nie utrzymuje się jeszcze między żadnym z jej stanów percepcyjnych a naparstkiem, nawet jeśli w jakimś stanie jej układu wzrokowego mogą się znajdować informacje, które umożliwiłyby komuś (na przykład zewnętrznemu obserwatorowi badającemu stany jej kory wzrokowej) zlokalizowanie i rozpoznanie naparstka. Betsy musi „skoncentrować się" na naparstku, odseparować go jako „figurę" od „tła" i zidentyfikować go. Kiedy to już się stanie, Betsy naprawdę widzi naparstek. Znalazł się on w końcu „w jej świadomym przeżyciu" – a teraz, gdy jest go świadoma, będzie przynajmniej w stanie podnieść ręce w geście zwycięstwa – lub po cichu usiąść z innymi dziećmi, które wcześniej go dostrzegły[107].

Tego rodzaju związki sterowane informacją zwrotną, regulowane i celowe są wymagane do zaistnienia znajomości zasługującego na swe miano – która następnie może funkcjonować na przykład jako podstawa strategii. Kiedy już dostrzegam coś w tym silnym sensie, mogę „coś z tym zrobić" albo zrobić, *ponieważ* widzę to lub *gdy tylko* to widzę. Pojedyncze naparstki, gdy zostaną już zidentyfikowane, są zwykle wystarczająco nieskomplikowane, aby następnie je śledzić (oczywiście z wyjątkiem sytuacji, gdy znajdujesz się w magazynie naparstków podczas trzęsienia ziemi). W normalnym wypadku wysoki status uzyskany przez naparstek w systemie sterującym Betsy nie trwa tylko przemijającej chwili; naparstek pozostanie zlokalizowany przez Betsy podczas sięgania po niego, w czasie potrzebnym do upewnienia się co do jego poprawnej identyfikacji czy powtórnego sprawdzenia go (i jeszcze raz – jeśli istnieją podstawy do zwątpienia). Rzeczy, których jesteśmy z całą pewnością świadomi, to elementy obserwowane szczerze i bez pośpiechu, gromadzące się i integrujące owoce wielu ruchów sakkadowych, tworzące z czasem znajomość, choć jednocześnie przedmiot pozostaje w przestrzeni prywatnej. Jeśli przedmiot przemieszcza się jak motyl, podejmiemy działanie, by go unieruchomić, abyśmy „mogli na niego patrzeć", a jeśli jest dobrze zakamuflowany, musimy podjąć kroki – dosłownie, jeśli nie możemy go dotknąć – aby ustawić go przed kontrastowym tłem.

Nasza porażka w takim momencie może uniemożliwić nam dojrzenie obiektu, w ważnym i znajomym znaczeniu tego słowa[108]. Obserwatorzy ptaków często posiadają życiową listę wszystkich gatunków, które widzieli. Załóżmy, że ty i ja jesteśmy takimi obserwatorami i razem słyszymy ptaka śpiewającego pośród drzew nad naszymi głowami; spoglądam w górę i mówię: „Widzę go – a ty?". Spoglądasz dokładnie tam, gdzie patrzę ja, a jednak mówisz zgodnie z prawdą: „Nie. Nie widzę go". Ja mogę wpisać tego ptaka na moją listę; ty nie, mimo że możesz mieć ogromną pewność, iż jego obraz musiał kilkakrotnie pojawić się w dołkach środkowych siatkówek twoich oczu.

Cóż więc mamy powiedzieć? Czy naparstek był w jakiś sposób „obecny" w świadomości Betsy, zanim go dojrzała? Czy ptak był obecny w „tle" twojej świadomości, czy nie był obecny w ogóle? Sprowadzenie czegoś na pierwszy plan świadomości oznacza sprowadzenie tego na pozycję, gdzie można zdać z niego relację, lecz co jest wymagane, aby sprowadzić coś na drugi plan świadomości (a nie jedynie na drugi plan środowiska wizualnego)? Naparstek i ptak bez wątpienia były obecne w środowisku wizualnym – nie w tym tkwi problem. Być może nie wystarczy, żeby światło odbite od obiektu jedynie dotarło do oczu, ale jaki dalszy efekt musi mieć odbite światło – jakie spostrzeżenie musi mieć mózg – aby obiekt przeszedł z poziomu zaledwie nieświadomego odzewu na drugi plan świadomego przeżycia?

Aby rozwiązać te trudności „pierwszoosobowego punktu widzenia", należy go pominąć i zbadać, czego można się dowiedzieć z punktu widzenia osoby trzeciej. W rozdziałach 8–10 odkrywaliśmy model realizacji mowy oparty na procesie typu Pandemonium, w którym ostateczne połączenie się treści z wyrażeniami było kulminacją konkurencji, budowaniem, rozbijaniem i ponownym budowaniem koalicji. Treści, które brały udział w tym boju, ale nie dały rady utrzymać się w nim zbyt długo, mogą wysłać pewnego rodzaju jednostrzałowy efekt „balistyczny" falujący w systemie, lecz byłby on niemal *niemożliwy do zrelacjonowania*. Gdy zdarzenie nie trwa długo, dowolna próba zrelacjonowania go, jeśli zostanie rozpoczęta, zostanie zaniechana lub będzie błądzić poza kontrolą, nie mając nic, względem czego mogłaby się poprawić. Aby coś mogło zostać zrelacjonowane, musi istnieć możliwość wielokrotnej identyfikacji efektu. Widzimy rozwój owej możliwości do relacjonowania w szkoleniach różnego rodzaju, przypominających te, których aplikowanie planowaliśmy u pacjentów ze ślepowidzeniem: rezultaty ćwiczenia wrażliwości podniebienia u sommelierów, trenowanie ucha u muzyków itp. – lub prosty eksperyment z szarpnięciem struny opisany w rozdziale 3.

Weźmy na przykład instrukcje dawane początkującym stroicielom fortepianów. Mówi się im, aby słuchali „bicia", gdy uderzają w klawisz, który stroją z dźwiękiem wzorcowym. *Jakiego* bicia? Początkowo większość uczących się nie jest w stanie odróżnić w swoim przeżyciu słuchowym nic, co odpowiadałoby opisowi „bicia" – to, co słyszą, mogliby opisać jako rodzaj pozbawionego struktury złego brzmienia czy rozstrojenia. W końcu jednak, jeśli trening jest sukcesem, są w stanie wyizolować w swoim przeżyciu słuchowym zakłócające „bicie" oraz zauważyć, w jaki sposób wzorce bicia zmieniają się w odpowiedzi na obracane przez nich kluczem kołki. Mogą sprawnie nastroić fortepian przez dostrajanie bicia. Zwykle mówią – a wszyscy możemy potwierdzić to podobnymi zdarzeniami z naszego własnego doświadczenia – że w wyniku treningu *zmieniło się ich świadome przeżycie*. Dokładniej mówiąc, zostało one poszerzone: są teraz świadomi rzeczy, których wcześniej świadomi nie byli.

Oczywiście w pewnym sensie słyszeli bicie cały czas. To w końcu zakłócenia stanowią „niedostrojenie", którego z pewnością byli świadomi. Jednak wcześniej nie byli w stanie wykryć tych składników w swoim doświadczeniu i dlatego można by powiedzieć, że te właściwości *przyczyniły się* do przeżycia, ale nie były same w nim *obecne*. Funkcjonalny status takiego czynnika przed szkoleniem był taki sam jak zdarzeń zachodzących w ślepowidzeniu: badany nie jest w stanie zrelacjonować pojedynczych czynników ani też zastosować strategii w momencie ich wystąpienia, lecz efekty tego czynnika nadal mogą się ujawnić w zachowaniu badanego, na przykład w jego umiejętności odpowiadania na umiejętnie zadane pytania. Sugeruję, że to właśnie oznacza bycie na drugim planie przeżycia, w jego tle. Oczywiście nie jest wykluczone (jak widzieliśmy), że wzmocnione połączenie, którego rodzaj właśnie opisaliśmy u stroicieli fortepianów czy sommelierów, mogłoby zostać zbudowane u pacjentów ze ślepowidzeniem do momentu, w którym zadeklarowalibyśmy i bez wahania zaakceptowali, że *stali się* świadomi bodźców – nawet na pierwszym planie swojej świadomości – których istnienie wcześniej mogli jedynie zgadywać.

Nie tak szybko – mówi Otto. – Oto kolejny zarzut. Wyobrażasz sobie pacjenta ze ślepowidzeniem uczącego się nowych sposobów używania zdolności ślepowidzenia i być może dałoby mu to *rodzaj* świadomości zdarzeń zachodzących w jego ślepym polu, lecz coś zostało pominięte. Świadomość ta nie byłaby *świadomością wizualną*; nie byłoby to jak *widzenie*. Brakowałoby „jakości fenomenalnych" czy też *qualiów* świadomego widzenia, nawet jeśli osoba ze ślepowidzeniem byłaby w stanie wykonać te funkcjonalne czynności.

Być może tak, być może nie. Czym dokładnie są „jakości fenomenalne" czy *qualia*? (*Qualia* to po łacinie *jakości*; liczba pojedyncza to *quale*). Z początku wydają się szalenie

oczywiste – to jakości wyglądu, zapachu, brzmienia, czucia rzeczy – ale osobliwie zmieniają status lub znikają w głębszej analizie. W następnym rozdziale dotrzemy do tych podejrzanych przez filozoficzne gęstwiny, lecz najpierw powinniśmy lepiej przyjrzeć się niektórym własnościom *niebędącym* jakościami fenomenalnymi, ale łatwo mogącym zostać z nimi pomylonymi.

4. Proteza wzroku: czego poza informacjami jeszcze brakuje?

Czy pacjent Weiskrantza, D.B., *widział* ruch? Cóż, z pewnością go nie *słyszał* ani nie *czuł*. Jednak czy jest to widzenie? Czy ma „jakości fenomenalne" widzenia? Weiskrantz mówi:

Gdy wzrasta „wyrazistość" bodźca, pacjent może twierdzić, że nadal nie „widzi", ale teraz ma rodzaj „przeczucia", że coś tam jest. W niektórych przypadkach, jeśli wyrazistość zostaje zwiększona jeszcze bardziej, może dojść do sytuacji, w której pacjent mówi, że „widzi", ale doświadczenie to nie jest wierne. Na przykład D.B. „widzi" w odpowiedzi na energicznie poruszający się bodziec, ale nie widzi go jako spójnego, poruszającego się obiektu, a za to mówi o złożonych wzorcach „fal". Inni badani mówią o „ciemnych cieniach" ujawniających się, gdy jasność i kontrast zostają zwiększone do wysokich poziomów. [Weiskrantz 1988, s. 189]

Energicznie poruszający się bodziec nie jest postrzegany przez D.B. jako mający barwę czy kształt, ale co z tego? Jak udowodniliśmy w rozdziale 3, w eksperymencie z kartą trzymaną w peryferyjnym polu widzenia, możemy przecież widzieć kartę, nie będąc w stanie zidentyfikować ani jej barw, ani kształtów. To zwykłe widzenie, nie ślepowidzenie, zatem nie powinniśmy na tej podstawie chętnie negować przeżycia wizualnego osoby badanej.

Pytanie, czy ten anomalny sposób uzyskiwania informacji o widzianych rzeczach może być odmianą *widzenia*, można sformułować bardziej obrazowo, gdy spojrzymy na jeszcze bardziej radykalne odejście od normalnego widzenia. Projektuje się przyrządy protetyczne mające zapewnić „wzrok" niewidomym, a niektóre z nich rodzą właśnie takie zagadnienia. W 1972 roku Paul Bach-y-Rita opracował kilka przyrządów zawierających małe kamery wideo o bardzo niskiej rozdzielczości, które mogły zostać zamontowane na oprawkach do okularów. Sygnał o niskiej rozdzielczości pochodzący z tych kamer, siatka 16 na 16 lub 20 na 20 „czarno-białych" pikseli, została umieszczona na plecach lub brzuchu pacjenta w sieci elektrycznie lub mechanicznie wibrujących mrowień zwanych „receptorami dotykowymi".

Ryc. 11.3
Niewidomy pacjent z 16-przewodowym, przenośnym systemem elektrycznym. Kamera telewizyjna jest podłączona do tulei soczewek zamontowanych na oprawkach do okularów. Mały zwój kabli prowadzi do elektrycznego obwodu napędzającego bodźce (pacjent trzyma go w prawej ręce). Matrycę 256. koncentrycznych, srebrnych elektrod pacjent trzyma w lewej ręce.

Po zaledwie kilku godzinach ćwiczeń niewidomi pacjenci posługujący się tym urządzeniem byli w stanie nauczyć się interpretować wzorce mrowienia na swojej skórze, tak jak możemy interpretować litery kreślone na swojej skórze czyimś palcem. Rozdzielczość jest niska, ale mimo to badani mogli nauczyć się czytać znaki i identyfikować obiekty, a nawet ludzkie twarze, jak możemy się domyślić, *patrząc* na fotografię sygnału pojawiającego się na monitorze oscyloskopu.

Ryc. 11.4
Wygląd 400-pikselowej reprezentacji twarzy kobiety widziany na monitorze oscyloskopu. Badani są w stanie poprawnie zidentyfikować wzorce bodźców na tym poziomie złożoności.

Efektem było oczywiście świadome, percepcyjne przeżycie wytworzone sztucznie, ale skoro dane wejściowe były rozłożone na plecach lub brzuchach pacjentów, czy było to *widzenie*? Czy przeżycie to miało „jakości fenomenalne" widzenia, czy tylko wrażenia dotykowego?

Przypomnijmy sobie jeden z naszych eksperymentów z rozdziału 3. Dotykowy punkt

widzenia łatwo przenosi się na koniuszek ołówka, co pozwala na odczucie nim tekstury bez jednoczesnej świadomości wibracji ołówka w palcach. Nie powinno nas więc dziwić, że podobny, choć ekstremalny efekt był udziałem pacjentów Bach-y-Rity. Po krótkim okresie szkolenia świadomość mrowienia na skórze zniknęła; można by powiedzieć, że podkładka z pikseli stała się przezroczysta i ich punkt widzenia przeniósł się na punkt widzenia kamery zamontowanej z boku ich głów. Niesamowitą demonstracją siły tej zmiany było zachowanie przeżywającej osoby, której kamera miała zoom z przyciskiem sterującym (Bach-y-Rita 1972, s. 98–99). Siatka mrowiących drapaczy znajdowała się na plecach, a kamera była zamontowana z boku głowy. Kiedy eksperymentator bez ostrzeżenia dotknął przycisku sterującego zoomem, sprawiając, że obraz na plecach badanego nagle rozrósł się czy też „zbliżył", badany instynktownie odsunął się *do tyłu, podnosząc ręce, aby ochronić głowę*. Kolejna szokująca demonstracja przezroczystości mrowienia jest taka, że badani, którzy byli zaznajomieni z siatką na plecach, potrafili niemal natychmiast zaadaptować się do siatki przeniesionej na brzuch (tamże, s. 33). Bach-y-Rita zauważa, że mimo to nadal reagowali na swędzenie na plecach jak na coś, co należy podrapać – nie twierdzili, że to „widzą" – i byli w stanie bez problemu zająć się mrowieniem jako mrowieniem, gdy zostali o to poproszeni.

Obserwacje te są spektakularne, ale nierozstrzygające. Można by zwrócić uwagę, że w momencie, gdy korzystanie z wejścia urządzenia stało się ich drugą naturą, badani naprawdę widzieli, lub przeciwnie – że tylko niektóre z najbardziej podstawowych cech „funkcjonalnych" widzenia zostały oddane przez protezę. Co z innymi „jakościami fenomenalnymi" widzenia? Bach-y-Rita mówi o wyniku pokazania dwóm badanym, niewidomym studentom, po raz pierwszy w ich życiu, fotografii kobiet z magazynu „Playboy". Byli zawiedzeni – „choć obaj byli w stanie opisać wiele treści fotografii, przeżycie to nie miało żadnego komponentu emocjonalnego; nie spowodowało żadnych przyjemnych doznań. To bardzo zaniepokoiło tych dwóch młodych mężczyzn, którzy wiedzieli, że podobne zdjęcia miały komponent emocjonalny dla ich widzących kolegów" (tamże, s. 145).

Urządzenia protetyczne nie dały więc *wszystkich* efektów normalnego widzenia. Część ich wad musi wynikać z dużej różnicy w tempie przepływu informacji. Normalne widzenie informuje nas o właściwościach przestrzennych w naszym środowisku z ogromną prędkością i z niemalże dowolnym poziomem szczegółów, jakiego sobie życzymy. Nie jest zaskoczeniem, że informacje przestrzenne o niskiej rozdzielczości przesyłane do mózgu przez interfejs na skórze nie wywołały wszystkich reakcji wywoływanych u ludzi widzących normalnie, gdy ich systemy wzrokowe są zalewane danymi[109]. Ile przyjemności powinno dać normalnie widzącej osobie *patrzenie* na odpowiedniki o podobnie niskiej rozdzielczości – rzuć okiem na rycinę 11.4 – przedstawiające pięknych ludzi?

Nie jest jasne, ile zmieniłoby się, gdybyśmy w jakiś sposób byli w stanie polepszyć „szybkość transmisji w bodach"[110] protezy wzroku, aby uplasowała się ona na poziomie normalnego widzenia. Być może zwykłe zwiększenie ilości i tempa informacji, co jakoś dałoby mózgowi mapy bitowe o większej rozdzielczości, wystarczyłoby do wytworzenia tego brakującego zachwytu. A przynajmniej jego części. Ludzie ślepi od urodzenia byliby daleko w tyle za tymi, którzy niedawno stracili wzrok, ponieważ nie mają oni żadnego charakterystycznego dla widzenia połączenia, które bez wątpienia odgrywa istotną rolę w przyjemności, jaką czerpią widzący ludzie ze swoich przeżyć, *przypominających* im o wcześniejszych doświadczeniach wizualnych. Być może również część przyjemności, którą czerpiemy z tych przeżyć, jest efektem ubocznym pradawnych śladów wcześniejszej ekonomii naszych systemów nerwowych – napomknąłem o tej kwestii w rozdziale 7, a rozwinę ją w rozdziale następnym.

Te same rozważania dotyczą ślepowidzenia i wszelkich wyobrażalnych ulepszeń w umiejętnościach pacjentów ze ślepowidzeniem. W debatach o ślepowidzeniu zwykle ignoruje się, jak kiepska jest informacja, którą ci badani zbierają ze swoich ślepych pól. Jedną rzeczą jest móc odgadnąć, gdy cię o to zapytają, czy w twoim ślepym polu zaprezentowano właśnie kwadrat czy koło. Byłoby czymś zupełnie innym potrafić odgadnąć ze szczegółami, gdy cię o to poproszą, co się właśnie działo za oknem.

To, czego dowiedzieliśmy się o widzeniu protetycznym, możemy wykorzystać do wyobrażenia sobie tego, czym byłoby dla pacjenta ze ślepowidzeniem odzyskanie większej ilości *funkcji* widzenia. Wyobraźmy sobie spotkanie z osobą cierpiącą na ślepotę korową, która po pilnym szkoleniu (1) uczyniła umiejętność zgadywania, kiedy zgadywać, swoją drugą naturą, (2) potrafi grać w chowanie naparstka z najlepszymi, (3) w jakiś sposób zdołała wielokrotnie zwiększyć prędkość i ilość szczegółów swojego działania związanego ze zgadywaniem. Spotykamy ją, gdy czyta gazetę i chichocze z komiksów, i prosimy o wyjaśnienia. Oto trzy scenariusze, ułożone w kolejności wzrastającej wiarygodności:

(1) Oczywiście, że tylko zgaduję! Nie widzę absolutnie nic, ale nauczyłam się, jak zgadywać, gdy należy zgadywać, i teraz na przykład zgaduję, że pokazujesz mi nieuprzejmy gest i wykrzywiasz twarz w ogromnym zdziwieniu.

(2) Cóż, to, co zaczęło się jako zwykłe zgadywanie, stopniowo przestało nim być i zaczęłam temu wierzyć. Stało się, powiedzmy, *przeczuciami*. Zwykle po prostu *wiedziałam*, że coś działo się w moim ślepym polu. Mogłam wtedy wyrazić tę wiedzę i postępować zgodnie z nią. Co więcej, miałam wówczas metawiedzę o tym, że w rzeczywistości mogłam mieć takie przeczucia, i potrafiłam użyć tej metawiedzy do planowania moich czynności i układania dla siebie strategii. To, co zaczęło się jako świadome zgadywanie, zamieniło się w świadome przeczucia, a teraz pojawiają się one tak szybko, że nie mogę ich nawet oddzielić. Ale nadal nic nie widzę! Nie tak jak widziałam kiedyś! To właściwie nie jest widzenie.

(3) No cóż, tak naprawdę to jest *bardzo podobne* do widzenia. Bez wysiłku podejmuję teraz czynności w świecie na podstawie informacji zebranych z otoczenia przez moje oczy. A jeżeli chcę, mogę być świadoma informacji napływających z oczu. Bez żadnego wahania reaguję na barwy obiektów, na ich kształty i lokalizację; straciłam całe poczucie wysiłku, który włożyłam w rozwinięcie tych talentów i sprawienie, że stały się moją drugą naturą.

A mimo to nadal moglibyśmy wyobrazić sobie naszą osobę badaną mówiącą o tym, że czegoś brakuje:

– *Qualia*. Moje stany percepcyjne oczywiście mają *qualia*, ponieważ są stanami świadomymi, ale zanim straciłam wzrok, miały one również *qualia wzrokowe*, a teraz ich nie mają, pomimo całego mojego szkolenia.

Może wydawać się oczywiste, że ma to sens, że jest to właśnie coś, czego moglibyśmy się spodziewać, iż powie nasza osoba badana. Jeśli tak, reszta tego rozdziału jest dla ciebie – ćwiczenie zaprojektowano tak, aby zburzyć to przekonanie. Jeśli zaczynasz już wątpić, że mowa o *qualiach* ma jakikolwiek sens, prawdopodobnie przewidujesz niektóre zwroty, które już niedługo nastąpią w naszej opowieści.

5. „Wypełnianie" kontra dowiadywanie się

Jednakże to, że istnieje takie poczucie obcości, nie jest jeszcze powodem, by mówić:

każdy dobrze znany przedmiot, gdy nie zdaje się nam obcy, daje nam poczucie swojskości. – Mniemamy niejako, iż miejsce, które zajmowało poczucie obcości, musi być przecież jakoś wypełnione.

Ludwig Wittgenstein, 1953/2000, s. 596

(przeł. Bogusław Wolniewicz)

W rozdziale 2 widzieliśmy, że jednym z powodów wiary w dualizm jest to, iż obiecuje on „materię, z której utkane są sny" – fioletowe krowy i inne wytwory naszej wyobraźni. W rozdziale 5 obserwowaliśmy, jakie nieporozumienia wyrosły z naturalnego, choć błędnego założenia, że *po tym*, jak mózg dokonał rozróżnienia lub oceny, ponownie *prezentuje* materiał, na którym ta ocena została oparta, dla przyjemności publiczności w teatrze kartezjańskim wypełniając go barwami. Idea *wypełniania* jest powszechna w myślach nawet wytrawnych teoretyków i jest to oczywista oznaka szczątkowego materializmu kartezjańskiego. To zabawne, ponieważ ci, którzy korzystają z tego terminu, są mądrzejsi, ale okazuje się on dla nich nie do odparcia, więc kryją się, umieszczając go w cudzysłowie.

Niemalże każdy na przykład opisuje mózg jako „wypełniający" plamkę ślepą (moje podkreślenie w każdym z przykładów):

[...] neurologicznie dobrze znane zjawisko subiektywnego „*wypełniania*" brakującej części ślepego pola w polu widzenia. [Libet 1985b, s. 567]

[...] możesz zlokalizować swoją własną plamkę ślepą, jak również zademonstrować, jak wzorzec zostaje w niej „*wypełniony*" lub „uzupełniony" [...]. [Hundert 1987, s. 427]

Mamy też słuchowe „wypełnianie". Gdy słuchamy mowy, luki w sygnale akustycznym mogą zostać „wypełnione" – na przykład w „efekcie przywracania fonemu" (Warren 1970). Ray Jackendoff ujmuje to w następujący sposób:

Pomyślmy na przykład o percepcji mowy z hałaśliwymi i wadliwymi danymi wyjściowymi – przypuśćmy w obecności poruszającego się samolotu czy przy kiepskim połączeniu telefonicznym. [...] To, co konstruujemy [...] to nie tylko zamierzone znaczenie, ale również struktura fonologiczna: „słyszymy" więcej, niż w rzeczywistości przekazuje nam dźwięk. [...] Innymi słowy, informacja fonetyczna zostaje „*wypełniona*" za pomocą struktur wyższego rzędu, jak również sygnału akustycznego; i mimo różnicy w tym, skąd pochodzi, nie ma różnicy jakościowej w samym uzupełnionym zdaniu. [Jackendoff 1987, s. 99]

A gdy czytamy tekst, zachodzi coś podobnego (ale wizualnego): jak ujmuje to Bernard Baars:

Odnajdujemy podobne zjawisko w dobrze znanym „efekcie robienia korekty", ogólnym wniosku, że błędy literowe w momencie wykonywania korekty są trudne do wykrycia, gdyż umysł „*wypełnia*" je poprawnymi informacjami. [Baars 1988, s. 173]

Howard Margolis daje niekontrowersyjny komentarz do całej kwestii „wypełniania":

„*Wypełnione*" szczegóły są po prostu poprawne. [Margolis 1987, s. 41]

Milczące rozpoznanie, iż jest coś podejrzanego w idei „wypełniania", przewija się w poniższym opisie plamki ślepej autorstwa filozofa C.L. Hardina w jego książce *Color for Philosophers*:

[Plamka ślepa] pokrywa obszar w polu widzenia o średnicy 6 stopni kątowych, co wystarczyłoby do ustawienia dziesięciu księżyców w pełni jeden przy drugim, a mimo to nie ma żadnej luki w odpowiadającym jej regionie pola widzenia. Dzieje się tak, gdyż oko-mózg *wypełnia* ją tym, co widziane jest w przylegających do niej obszarach. Jeśli jest tam niebieski, luka zostaje *wypełniona* na niebiesko; jeśli jest tam szkocka krata, nie jesteśmy świadomi

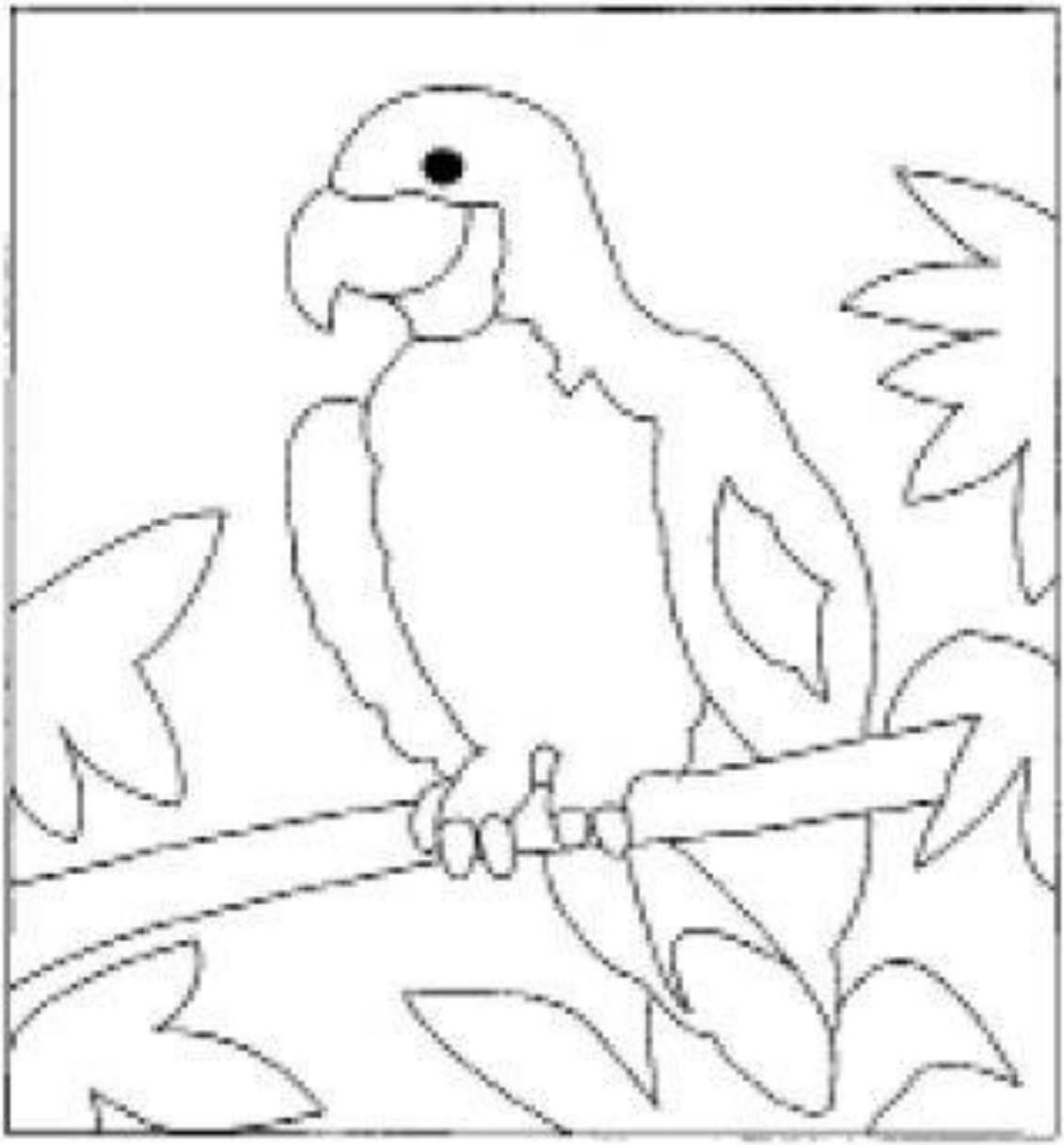
żadnego braku ciągłości we wzorze kraty. [Hardin 1988, s. 22]

Hardin nie potrafi się zmusić do powiedzenia, że mózg wypełnia kratę, gdyż to z pewnością sugerowałoby dosyć wyszukaną „konstrukcję", niczym fantazyjne „niewidoczne łatanie", za które możesz sporo zapłacić, gdy chcesz, aby wypełniono twoją marynarkę w jodełkę: wówczas wszystkie linie się ze sobą zgadzają i wszystkie odcienie barw na granicy są do siebie dopasowane. Wydaje się, że wypełnienie niebieskim to jedna kwestia – wystarczyłyby ze dwa maźnięcia mózgową farbą odpowiedniej barwy; jednak wypełnienie szkockiej kraty to co innego, i jest to więcej, niż jest on w stanie uznać.

Ale komentarz Hardina przypomina nam, że jesteśmy tak samo nieświadomi naszych plamek ślepych, kiedy patrzymy na kratę, jak wówczas, gdy spoglądamy na jednobarwną powierzchnię, zatem to, co powoduje naszą nieświadomość, może być w taki sam sposób uzyskane przez mózg w obu przypadkach. „Nie jesteśmy świadomi żadnego braku ciągłości", mówi. Jeśli jednak mózg nie musi wypełniać luki kratą, dlaczego miałby chcieć wypełnić lukę niebieskim?

W żadnym z tych przypadków „wypełnianie" nie jest prawdopodobnie kwestią dosłownego wypełniania – w rodzaju tego, które wymaga użycia pędzli. (Był to morał historii o CAD Dla Niewidomych 2.0 w rozdziale 10). Nie zakładam, że ktoś uważa, iż „wypełnianie" sprowadza się do tego, że mózg rzeczywiście zajmuje się malowaniem jakiejś przestrzeni *pigmentem*. Wiemy, że prawdziwy, odwrócony do góry nogami obraz na siatkówce jest ostatnim etapem widzenia, w którym cokolwiek jest kolorowane w sposób równie bezproblemowy, jak kolorowanie filmowego obrazu na ekranie. Skoro nie istnieją dosłowne oczy wyobraźni, nie ma pożytku z pigmentu w mózgu.

To by było na tyle z pigmentem. Nadal jednak może nam się wydawać, że w mózgu następuje coś, co jest *jakoś* istotnie analogiczne do zamalowania powierzchni pigmentem – w przeciwnym razie w ogóle nie chcielibyśmy mówić o „wypełnianiu". Czymkolwiek jest, najwyraźniej to coś specjalnego w przestrzennym „medium" przeżycia wizualnego czy audialnego. Jak mówi Jackendoff, mając na myśli przypadek słuchowy, „»słyszymy« więcej, niż sygnał w rzeczywistości przekazuje" – ale zauważmy, że nadal umieszcza wyraz „słyszymy" w cudzysłowie. Cóż takiego może być *obecne*, gdy „słyszymy" dźwięki wypełniające ciche momenty lub „widzimy" barwy wypełniające puste przestrzenie? Rzeczywiście wydaje się, że coś tam jest, coś, co mózg musi dostarczyć (poprzez „wypełnienie"). Jak powinniśmy to nazwać, czymkolwiek jest? Nazwijmy to *wymysłem*. Kusi nas więc twierdzenie, że istnieje coś zrobione z wymysłu, co pojawia się, gdy mózg „wypełnia", a czego nie ma, gdy nie kłopocze się „wypełnianiem". Tak ujęty pomysł wypełnienia oczywiście nie spodoba się wielu osobom. (Przynajmniej mam taką nadzieję). Jesteśmy mądrzejsi: nie istnieje substancja taka jak wymysł. Mózg nie tworzy wymysłu; mózg nie używa wymysłu do wypełnienia luk; wymysł jest jedynie wymysłem mojej wyobraźni. To by było na tyle z wymysłem! Jednak co znaczy „wypełnianie", co *mogłoby* znaczyć, jeśli nie wypełnianie wymysłem? Jeśli nie istnieje medium w postaci wymysłu, czym różni się „wypełnianie" od niekłopotania się wypełnianiem?



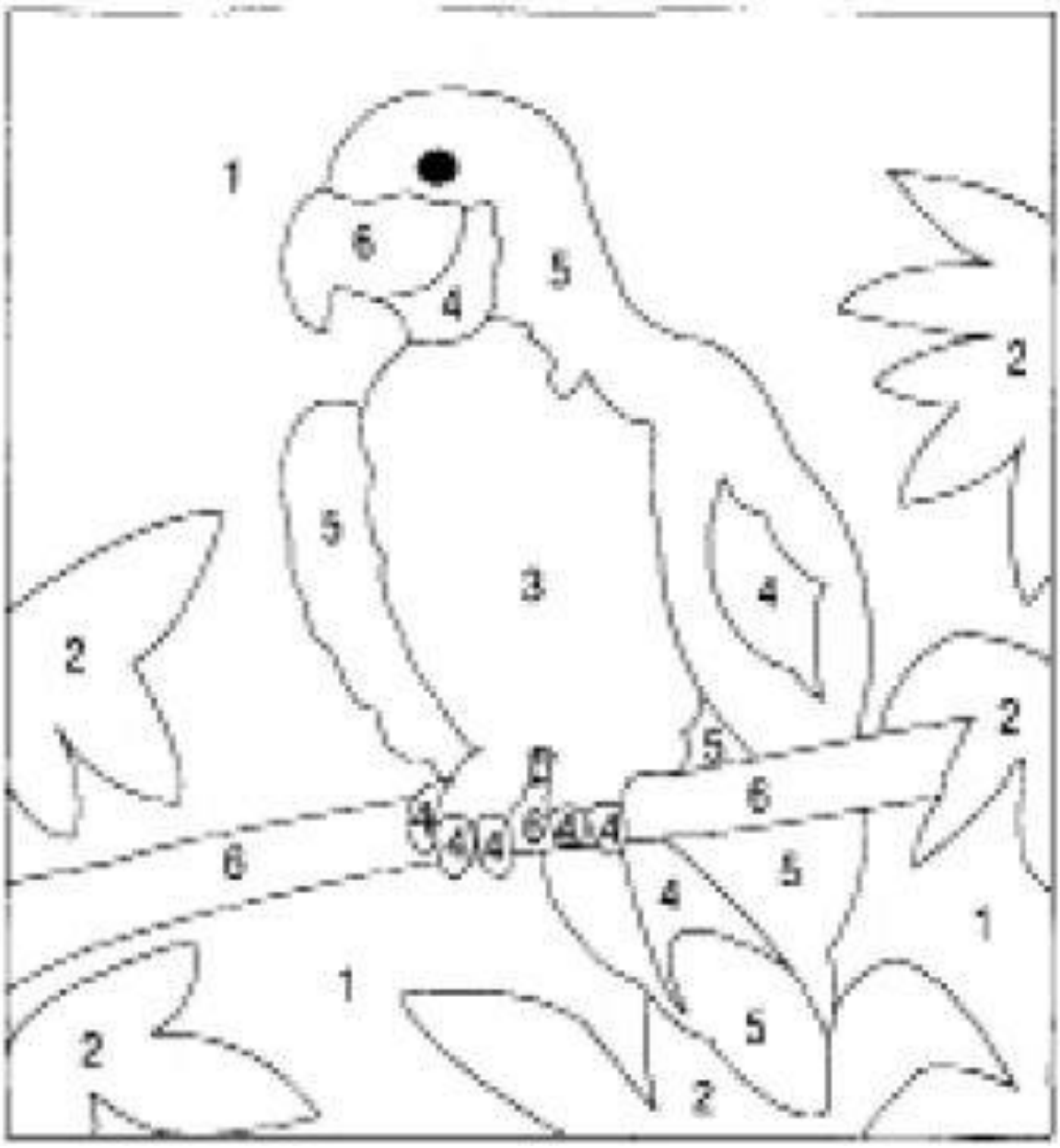
Ryc. 11.5

W rozdziale 10 widzieliśmy, że system CAD mógł reprezentować barwy, łącząc ich numer z każdym pikselem lub z ograniczonym regionem przedstawianego obiektu. Widzieliśmy też, jak system CAD Dla Niewidomych 2.0 mógł szukać czy wykrywać barwy, sczytując taki kod. Proces ten przypomina dziecięcą kolorowankę z liczbami, która jest prostą analogią procesów reprezentacji muszących lub mogących następować w mózgu. Rycina 11.5 jest reprezentacją z informacjami o kształtach, ale bez informacji o kolorach.

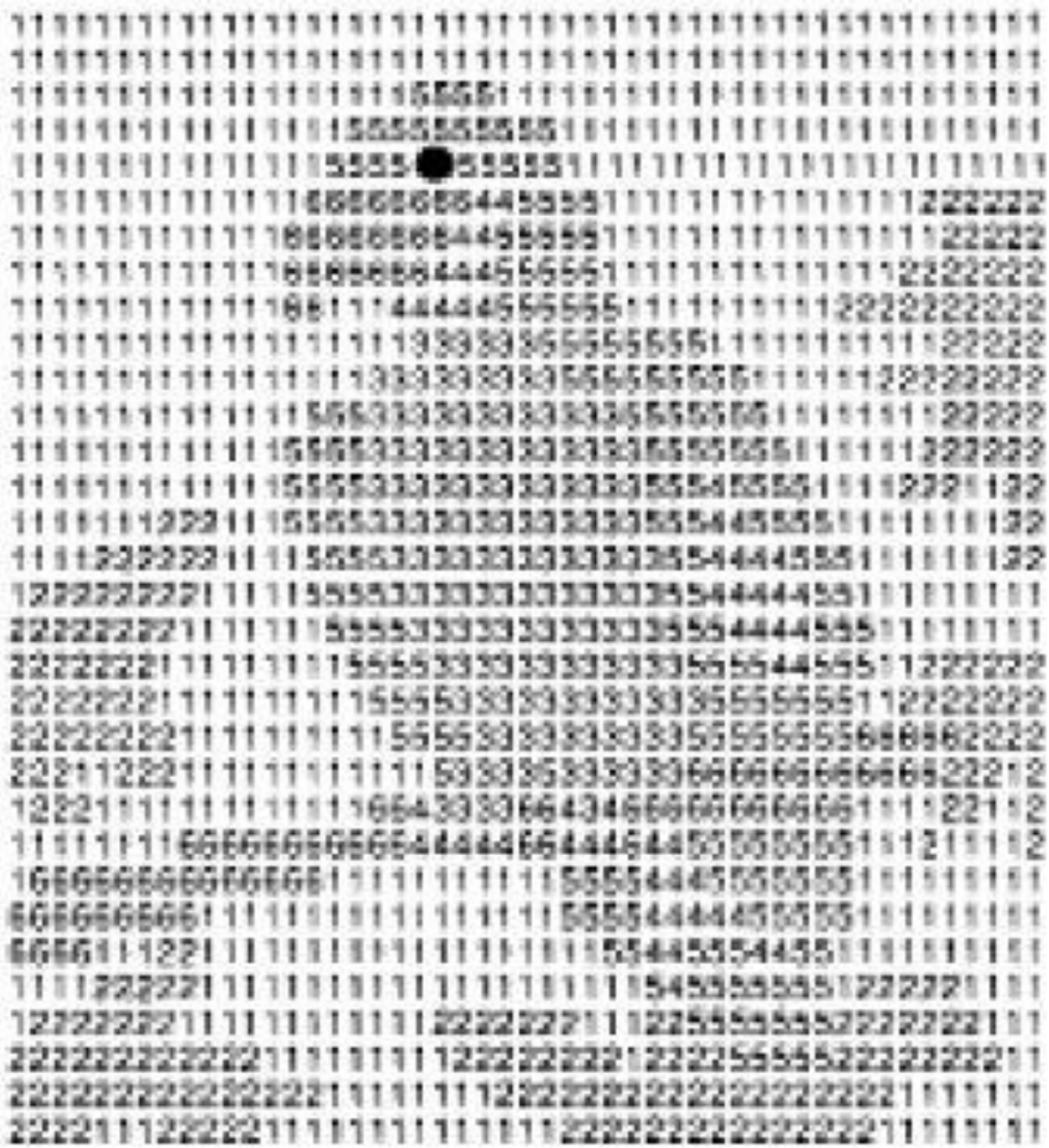
Porównajmy ją z ryciną 11.6, która zawiera informacje o kolorach w postaci kodu liczbowego. Kredkami można by wypełnić obszary kolorami zgodnie z poleceniem, przemieniając rycinę 11.6 w kolejny rodzaj „wypełnionej" reprezentacji – takiej, w której obszary wypełnione są prawdziwym kolorem, prawdziwym pigmentem.

Jest jeszcze jeden sposób wypełnienia kolorami, piksel po pikselu, czyli zakodowana mapa bitowa, przedstawiona na rycinie 11.7.

Ryciny 11.6 i 11.7 są rodzajami wypełniania (w porównaniu na przykład z ryciną 11.5), ponieważ każda procedura, która potrzebuje informacji o kolorze obszaru, może, poprzez jego mechaniczne zbadanie, tę informację wydobyć. Jest to wypełnianie czysto informacyjne. Te systemy są oczywiście całkowicie arbitralne. Możemy bez problemu stworzyć nieskończoną liczbę funkcjonalnie równoważnych systemów reprezentacji – korzystających z innych systemów kodowania czy innych mediów.



Ryc. 11.6
1 – niebieski, 2 – zielony, 3 – pomarańczowy, 4 – czerwony, 5 – fioletowy, 6 – żółty



Ryc. 11.7

Jeśli tworzysz kolorowy obrazek na swoim komputerze w odpowiednim programie, widziany przez ciebie ekran jest przez maszynę reprezentowany jako mapa bitowa w „buforze ramek", analogiczna do ryciny 11.7, jednak gdy zapisujesz obrazek na dysku, algorytm kompresji tłumaczy go na coś podobnego do ryciny 11.6. Dzieli on cały obszar na regiony o podobnym kolorze i przechowuje granice regionów oraz numery ich kolorów w pliku „archiwum"[111]. Jest on równie dokładną mapą bitową, lecz wprowadzając generalizację dotyczącą regionów i oznakowując każdy z nich tylko raz, jest bardziej wydajnym systemem reprezentacji.

Mapa bitowa, dokładnie oznakowując każdy piksel, jest rodzajem czegoś, co moglibyśmy nazwać reprezentacją *„z grubsza ciągłą"* – jej ziarnistość jest funkcją rozmiaru pikseli. Mapa

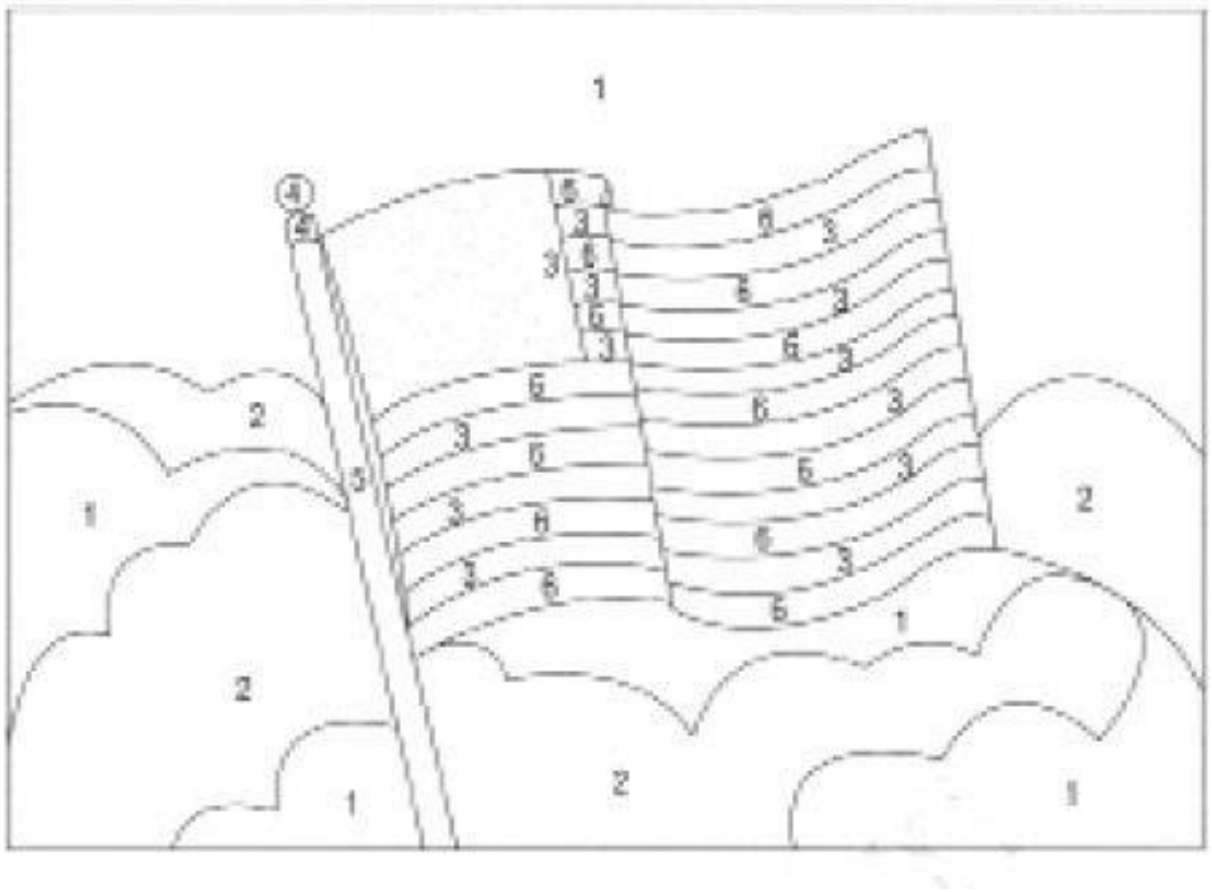
bitowa nie jest dosłownie obrazem, a jedynie tablicą wartości, swego rodzaju przepisem na obrazek. Tablica może być zapisana w każdym systemie przechowującym informacje o lokalizacji. Kaseta wideo to kolejny nośnik z grubsza ciągłej reprezentacji, jednak to, co jest na niej przechowywane, to nie dosłowne obrazy, ale przepisy (na różnych poziomach ziarnistości) na stworzenie obrazów.

Kolejnym sposobem zapisu obrazu na ekranie komputera jest przechowanie kolorowej fotografii na, powiedzmy, 35-milimetrowym slajdzie, lecz w sposób oczywisty różni się to znacznie od innych systemów: istnieje prawdziwa farba dosłownie wypełniająca obszar prawdziwej przestrzeni. Jak w przypadku mapy bitowej, jest to z grubsza ciągła reprezentacja przedstawionych obszarów przestrzeni (ciągła aż do najmniejszego ziarna filmu – w odpowiedniej skali staje się jak piksele, czyli ziarnista). Jednak w przeciwieństwie do mapy bitowej, do reprezentowania koloru użyty jest kolor. Kolorowy negatyw również używa koloru do reprezentacji koloru, ale w sposób odwrócony.

Mamy zatem trzy sposoby „wypełniania" informacji kolorystycznych: kolory zgodne z liczbami, jak na rycinie 11.6 i w pliku archiwum, kolory zgodne z mapą bitową, jak na rycinie 11.7 i w buforze ramek czy na kasecie wideo, oraz kolory zgodne z kolorami. Kolory zgodne z liczbami to w pewnym sensie „wypełnianie" informacji kolorystycznych, jednak w porównaniu z innymi systemami jest wydajne właśnie dlatego, że *nie* kłopocze się wypełnianiem wartości dla każdego piksela z osobna. Na który z tych sposobów (jeśli taki został tu opisany) mózg „wypełnia" plamkę ślepą? Nikt nie uważa, że mózg używa liczb, pod którymi kryją się zakodowane kolory, ale nie w tym tkwi problem. Taki rejestr liczb może być rozumiany jako odpowiednik jakiegokolwiek systemu wielkości, jakiegokolwiek systemu „wektorów", który mózg mógłby zastosować jako „kod" kolorów; mogłaby to być częstotliwość wyładowań neuronalnych lub jakiś system adresów czy lokalizacji w sieciach neuronowych, czy też jakikolwiek inny system zmiennych fizycznych w mózgu. Rejestr numerów ma sympatyczną cechę: zachowuje relacje między wielkościami fizycznymi, jednocześnie pozostając neutralny w kwestii „wewnętrznych" właściwości takich wielkości, więc mogą one odpowiadać wszelkim wielkościom fizycznym w mózgu „kodującym" kolory. Mimo że liczby mogą być używane w sposób zupełnie arbitralny, mogą również być użyte niearbitralnie, aby odzwierciedlić strukturalne relacje między odkrytymi kolorami. Popularna „bryła barw", w której odcień, nasycenie oraz jasność to trzy wymiary, jakimi barwy różnią się od siebie[112], jest logiczną przestrzenią idealnie nadającą się do numerycznego potraktowania – każdego traktowania numerycznego odzwierciedlającego relacje znajdowania się pomiędzy, relacje przeciwieństw i dopełnień itp., które wykazuje ludzki wzrok. Im więcej wiemy o tym, jak mózg koduje barwy, tym potężniejszy i niearbitralny numeryczny model ludzkiego widzenia barw będziemy w stanie stworzyć.

Kłopot z mówieniem o „kodowaniu" barw przez mózg poprzez natężenia lub wielkości takiej czy innej rzeczy jest taki, że sugeruje nieostrożnym, iż koniec końców te kody muszą zostać odczytane, „powracając" do barwy. Jest to jednak trasa – być może najbardziej popularna – z powrotem do wymysłu: wyobrażamy sobie, że mózg mógłby nieświadomie przechowywać swoje encyklopedyczne informacje o barwie w formacie takim jak ten na rycinie 11.8, lecz następnie przeprowadzać „dekodowanie" reprezentacji do „prawdziwych barw" podczas specjalnych okazji – jak puszczenie kasety wideo, aby odzwierciedlić prawdziwe barwy na ekranie. Z pewnością jest różnica w fenomenologii między zaledwie przypominaniem sobie sugestii, że flaga jest czerwona, biała i niebieska, a rzeczywistym wyobrażaniem sobie „barwnej" flagi oraz „widzeniem" (oczyma wyobraźni), że jest ona czerwona, biała i niebieska. Jeśli ten kontrast w fenomenologii zachęca niektórych do twierdzenia, że istnieje wymysł, to jeszcze

bardziej frapującym przykładem jest złudzenie neonowej poświaty (van Tuijl 1975), którego przykład można zobaczyć na tylnej okładce tej książki.



Ryc. 11.8
1 – niebieski, 2 – szary, 3 – biały, 4 – złoty, 5 – czarny, 6 – czerwony

Róż, który widzisz jako wypełniający pierścień ograniczony czerwonymi liniami, nie jest wynikiem rozmazania tego koloru ani rozproszenia światła. Innymi słowy, nie ma barwy różowej na obrazie w siatkówce, poza czerwonymi linami. Jak można wyjaśnić tę iluzję? Jeden mózgowy obwód wyspecjalizowany w rozpoznawaniu kształtów zostaje wprowadzony w błąd i wyróżnia ograniczony obszar: pierścień z jego „subiektywnymi konturami". Subiektywne kontury są wytwarzane przez wiele podobnych rycin, takich jak te poniżej.

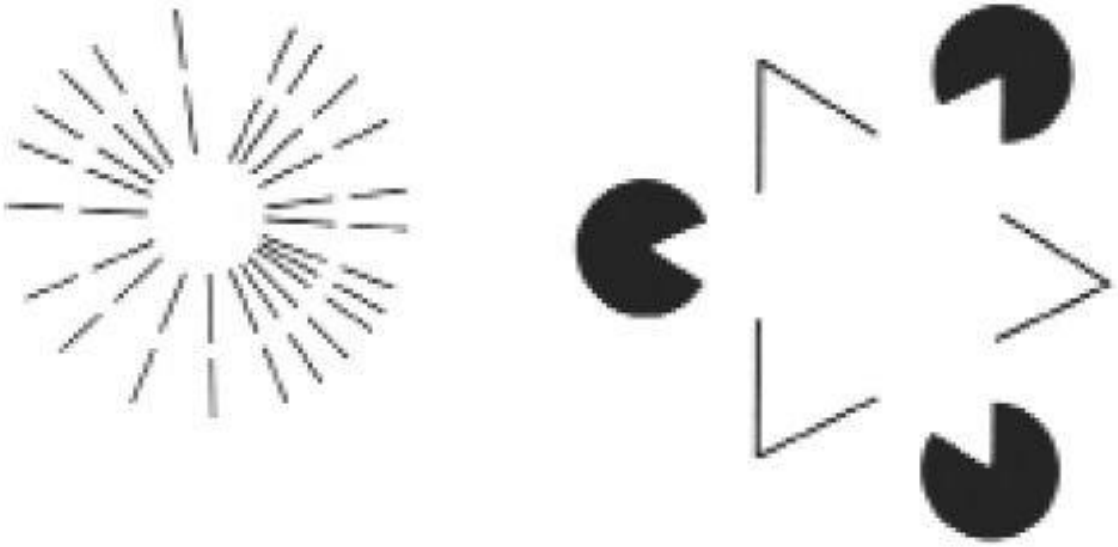


Рис. 11.9

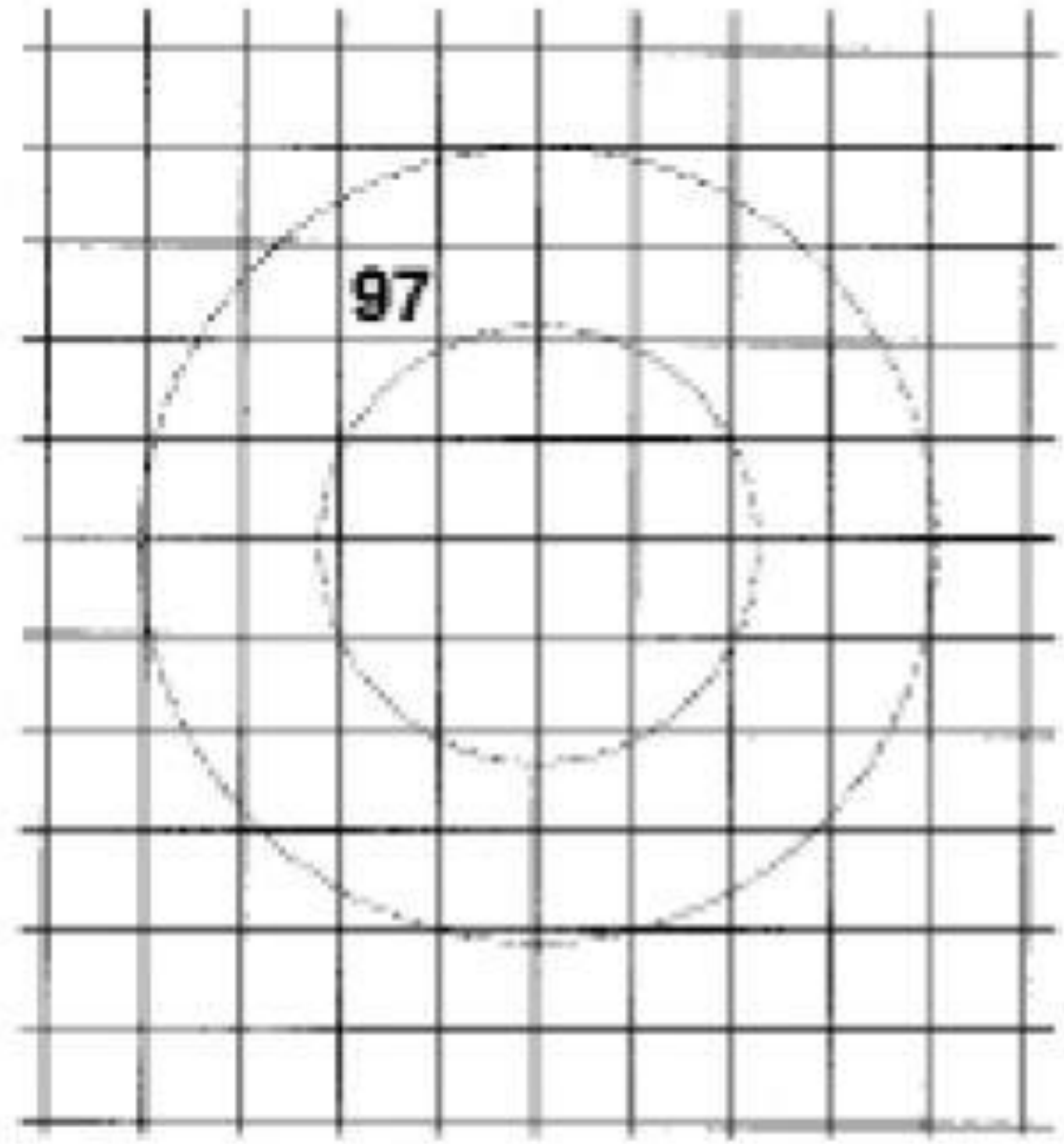
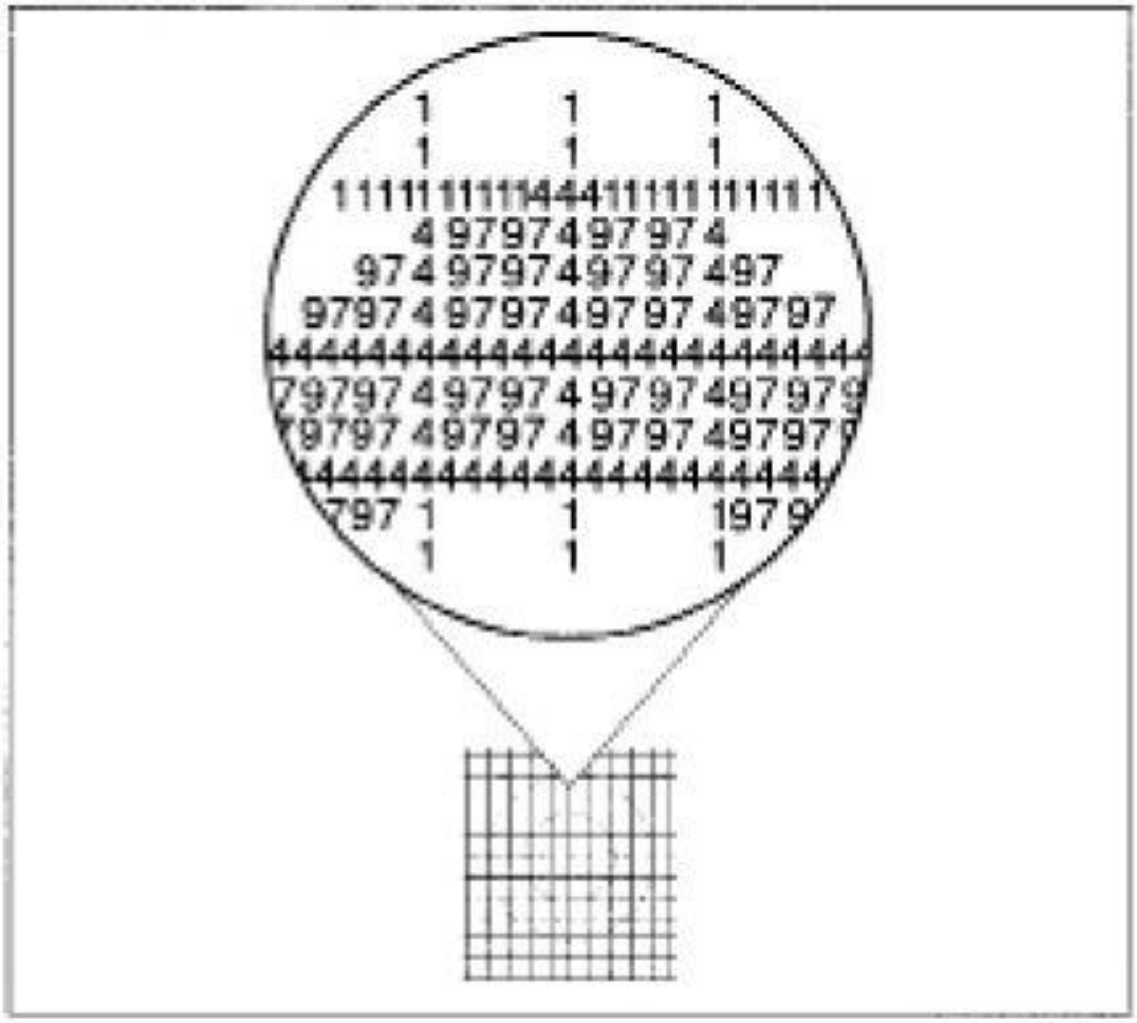


Рис. 11.10



Ryc. 11.11

Kolejny obwód mózgowy, specjalizujący się w barwie, ale niedający sobie rady z kształtem i lokalizowaniem, tworzy rozróżnienie barwy (powiedzmy róż nr 97), którym oznakowuje coś w sąsiedztwie, a etykieta zostaje przyczepiona (czy też „przypisana") całemu obszarowi.

To, dlaczego te konkretne rozróżnienia następują w tych warunkach, nadal pozostaje sporne, ale spór dotyczy mechanizmów przyczynowych, prowadzących do błędnego oznakowania obszaru, a nie dalszych „wymysłów" (jeśli takowe w ogóle istnieją) systemu wizualnego. Czy czegoś jednak nie brakuje? Poprzestałem na wyjaśnieniu dostarczającym regiony oznakowane liczbami odpowiadającymi barwom: Czy ten przepis na obraz barwny nie musi być gdzieś wykonany? Czy róż nr 97 nie musi być „wypełniony"? Możesz w końcu czuć pokusę, by stwierdzić, że widzisz róż! Z pewnością nie widzisz ograniczonego obszaru z wpisanym weń numerem. Dostrzegany przez ciebie róż nie istnieje w świecie zewnętrznym (nie jest to pigment, farba czy „zabarwione światło"), więc musi być „tutaj" – innymi słowy, różowy wymysł.

Musimy być ostrożni, aby odróżnić hipotezę „różowego wymysłu" od innych ujęć potencjalnie konkurencyjnych wobec wyjaśnienia, które poprzestało na sugestii oznakowania

barw liczbami. Może się na przykład okazać, że gdzieś w mózgu istnieje z grubsza nieprzerwana reprezentacja obszarów barwnych – mapa bitowa – w której „każdy piksel" w obszarze musi być oznakowany „barwa nr 97", mniej więcej jak na rycinie 11.11.

Jest to możliwość empiryczna. Moglibyśmy zaprojektować doświadczenia ją potwierdzające lub obalające. Pytanie brzmiałoby: Czy istnieje medium reprezentacyjne w mózgu, w którym wartość jakiegoś zmiennego parametru (intensywność bądź cokolwiek kodowane jako barwa) musi zostać rozpropagowana lub powielona przez odpowiednie piksele w matrycy, czy może istnieje po prostu „pojedyncze oznakowanie" obszaru, bez wymagania dalszego „wypełniania" lub „oświecania"? Jakie rodzaje eksperymentów mogłyby wesprzeć taki model neonowej poświaty? Cóż, byłoby na przykład godne podziwu, gdyby barwa w jakichś warunkach mogła zostać pokazana jako rozprzestrzeniająca się wolno w czasie – rozpoczynając od centralnych, czerwonych linii i stopniowo docierając do subiektywnych granic konturów[113]. Nie chcę z góry rozstrzygać pytania, gdyż moim głównym celem jest zilustrowanie twierdzenia, że choć jest mnóstwo nierozstrzygniętych pytań empirycznych dotyczących tego, jak zjawisko neonowej poświaty zachodzi w mózgu, żadne z nich nie dotyczy sporów o to, czy wymysł jest tworzony poprzez „dekodowanie" w systemie kodowania neuronowego, czy też nie.

Pytanie o to, czy mózg „wypełnia" w taki czy inny sposób, nie jest kwestią, którą może rozstrzygnąć jedynie introspekcja, gdyż, jak widzieliśmy w rozdziale 4, introspekcja dostarcza nam – osobie badanej i „zewnętrznemu" eksperymentatorowi – jedynie treść reprezentacji, a nie cechy reprezentującego nośnika. Aby uzyskać świadectwa związane z tym nośnikiem, musimy przeprowadzić kolejne eksperymenty[114]. Jednak w przypadku niektórych zjawisk możemy być *raczej* pewni, że nośnikiem reprezentacji jest coś wydajnego, jak kolory kodowane liczbami, nie z grubsza ciągłego, jak mapy bitowe.

Pomyślmy na przykład, jak mózg radzi sobie z tapetą. Załóżmy, że wchodzisz do pokoju i zwracasz uwagę, że tapeta to regularna siatka setek identycznych łódek lub – oddajmy hołd Andy'emu Warholowi – identyczne, fotograficzne portrety Marilyn Monroe. Aby zidentyfikować obrazek jako portret Marilyn Monroe, musisz dopuścić go do dołka środkowego siatkówki oka: obraz musi znaleźć się w wysokiej rozdzielczości dołka środkowego w oczach. Jak było widać w eksperymencie z kartą w rozdziale 2, widzenie peryferyjne (za które odpowiedzialna jest reszta siatkówki) nie ma zbyt wysokiej rozdzielczości; nie możesz nawet zidentyfikować waleta karo trzymanego w wyciągniętej ręce. Wiemy jednak, że po wejściu do pokoju, na którego ścianach znajdują się identyczne zdjęcia Marilyn Monroe, „natychmiast" sobie to uświadomisz. W ciągu ułamka sekundy zauważysz, że jest tam „mnóstwo identycznych, szczegółowych, wyraźnych portretów Marilyn Monroe". Ponieważ oczy wykonują co najwyżej cztery albo pięć ruchów sakkadowych na sekundę, możesz ustawić w dołku środkowym jedynie jedną lub dwie Marilyn w czasie potrzebnym do wyciągnięcia wniosku i *zaraz potem zobaczenia* setek identycznych Marilyn. Wiemy, że widzenie peryferyjne *nie mogłoby* odróżnić Marilyn od różnych plam w kształcie Marilyn, ale mimo to tym, co widzisz, *nie* jest tapeta z Marilyn pośrodku, otoczona różnymi plamami w kształcie Marilyn.

Czy jest możliwe, że mózg bierze jeden z obrazów Marilyn o wysokiej rozdzielczości z dołka środkowego i powiela go, jak w ksero, na wewnętrznym odbiciu ściany? Tylko tak szczegóły o wysokiej rozdzielczości, użyte do zidentyfikowania Marilyn, mogą w ogóle „dostać się na *drugi* plan", skoro widzenie peryferyjne nie jest wystarczająco ostre, aby to zapewnić. Przypuszczam, że zasadniczo jest to możliwe, lecz mózg prawie na pewno nie kłopocze się *tym* wypełnianiem! Gdy już zidentyfikuje jedną Marilyn i nie otrzyma żadnej informacji o tym, że inne plamy nie są Marilyn, dochodzi do wniosku, iż reszta to Marilyn, i oznacza cały obszar jako „więcej Marilyn", bez żadnego dalszego rozpoznania Marilyn[115].

Oczywiście ty tego tak nie postrzegasz. Tobie wydaje się, że rzeczywiście widzisz setki identycznych Marilyn. I w pewnym sensie masz rację: rzeczywiście na ścianie są setki identycznych Marilyn, a ty je widzisz. Nie jest jednak prawdą, że setki identycznych Marilyn są reprezentowane w mózgu. Twój mózg w jakiś sposób reprezentuje *to, że* są tam setki identycznych Marilyn, i bez względu na to, jak żywe jest twoje wrażenie, że widzisz wszystkie te szczegóły, są one w świecie, a nie w głowie. I nie zostaje użyty żaden wymysł, aby przedstawić to wydawanie się, gdyż nie jest ono w ogóle przedstawione, nawet jako mapa bitowa.

Teraz zatem możemy odpowiedzieć na nasze pytanie o plamkę ślepą. Mózg nie musi jej „wypełniać", ponieważ obszar, na którym wypada, jest już oznaczony (np. „kratka", „Marilyn", czy po prostu „więcej tego samego"). Gdyby mózg otrzymał sprzeczne świadectwa z jakiegoś obszaru, porzuciłby lub dostosowałby swoje generalizacje, ale brak świadectwa z plamki ślepej nie jest tym samym co otrzymanie sprzecznych świadectw. Brak potwierdzających świadectw z obszaru plamki ślepej nie jest problemem dla mózgu; mózg nie otrzymał wcześniej informacji z luki w siatkówce, więc nie rozwinął żadnego epistemologicznie głodnego ośrodka, który wymagałby nakarmienia informacjami z tego obszaru. Pośród wszystkich homunkulusów wzroku żaden nie ma roli koordynowania informacji z tego obszaru oka, więc gdy nie dociera żadna informacja z tych źródeł, nikt nie narzeka. Ten region jest po prostu zaniedbany. Innymi słowy, wszyscy normalnie widzący ludzie „cierpią" na malutką „anosognozję". Jesteśmy nieświadomi naszego „deficytu" – faktu, że nie otrzymujemy żadnej informacji wizualnej z naszych ślepych plamek. (Dobre przedstawienie anosognozji znajdziesz w McGlynn i Schacter 1989).

Plamka ślepa to dziura przestrzenna, ale mogą istnieć również dziury czasowe. Najmniejsze są luki pojawiające się, gdy nasze oczy skaczą sakkadowo. Nie zauważamy ich, lecz nie muszą one być wypełnione, *gdyż* jesteśmy skonstruowani tak, aby ich nie widzieć. Czasowymi analogami obszarów ślepych mogą być „nieobecności" podczas małych napadów padaczkowych. Są one zauważalne przez chorego, ale jedynie poprzez wnioskowanie: nie potrafi on „dostrzec granic", tak jak ty nie potrafisz dostrzec granic swojej plamki ślepej, jednak po napadzie może być zdziwiony nieciągłościami w zdarzeniach, które przeżył.

Podstawowa wada pomysłu „wypełniania" jest taka, że sugeruje, iż mózg czegoś dostarcza, gdy tak naprawdę mózg coś ignoruje. Prowadzi to nawet bardzo wyrafinowanych myślicieli do katastrofalnych błędów, świetnie streszczonych przez Edelmana (1989, s. 119): „Jedną z najbardziej szokujących właściwości świadomości jest jej ciągłość". Jest to całkowicie błędne. Jedną z najbardziej szokujących właściwości świadomości jest jej nieciągłość – jak pokazuje nam plamka ślepa czy luki sakkadowe, żeby podać najprostsze przykłady. Nieciągłość świadomości jest zaskakująca z powodu *pozornej* ciągłości świadomości. Naumann (1990) wskazuje na to, że świadomość może, ogólnie rzecz biorąc, być zjawiskiem z lukami i dopóki czasowe brzegi tych luk nie są postrzegane, nie będzie ich poczucia w „strumieniu" świadomości. Jak mówi Minsky: „Nic nie może *wydawać się* urywane, oprócz tego, co jest *reprezentowane* jako urywane. Paradoksalnie nasze poczucie ciągłości pochodzi od naszego wspaniałego *braku wrażliwości* na większość rodzajów zmian, a nie z jakiejś autentycznej spostrzegawczości" (Minsky 1985, s. 257).

6. Zaniedbanie jako patologiczna utrata apetytu episemologicznego

Motto mózgu dotyczące radzenia sobie z plamką ślepą mogłoby brzmieć: nie zadawaj mi pytań, a ja cię nie okłamię. Jak widzieliśmy w rozdziale 1, dopóki mózg zaspokaja każdy epistemologiczny głód, nie pozostaje nic, co musi robić. Ale co z momentami, kiedy jest o wiele mniej tego głodu, niż powinno być? Są to patologie zaniedbywania.

Jedną z jej najbardziej znanych form jest zespół jednostronnego zaniedbywania przestrzeni, w którym jedna strona ciała, zwykle lewa, zostaje całkowicie zaniedbana z powodu uszkodzenia przeciwnej półkuli mózgu. Nie tylko lewa strona ciała, ale również lewa strona najbliższego sąsiedztwa. Jeśli grupa ludzi staje wokół łóżka pacjenta z zaniedbywaniem prawostronnym, będzie on patrzył tylko na ludzi stojących po jego prawej stronie; gdy poprosi się go o policzenie ludzi znajdujących się w pomieszczaniu, zwykle pominie tych stojących z lewej strony, a gdy ktoś z lewej strony spróbuje zwrócić jego uwagę, zwykle nie będzie to miało żadnego skutku. Mimo to można pokazać, że organy zmysłowe pacjenta nadal przyjmują informacje, analizują je i reagują różnorako na bodźce po lewej stronie. Co może się dziać w głowie pacjenta? Czy „lewa strona przestrzeni fenomenalnej jest pusta"? Czy „oczyma wyobraźni" pacjent nie jest w stanie dostrzec materiału dostarczanego przez mózg po lewej stronie... sceny teatru kartezjańskiego?

Istnieje łatwiejsze wyjaśnienie, nie w kategoriach wewnętrznych reprezentacji o osobliwych właściwościach, ale w kategoriach *zaniedbania* w sensie politycznym! Wiadomo, że Daniel Patrick Moynihan uważał, iż pewne problemy w relacjach między rasami Ameryki rozwiązałyby się same, gdybyśmy traktowali je z „nieszkodliwym zaniedbaniem" – gdyby Waszyngton i reszta narodu po prostu zignorowała je na jakiś czas. Nie uważam, że była to dobra rada, ale Moynihan miał rację co do jednej rzeczy: istnieją warunki, w których nieszkodliwe zaniedbanie jest potrzebne – takie jak nasz problem plamki ślepej.

Nie ma homunkulusów, jak to ująłem, które miałyby „dbać" o informacje pochodzące z części pola widzenia odpowiadającego plamkom ślepym, więc gdy nic nie dociera, nikt nie marudzi. Być może różnica między nami a osobami cierpiącymi na patologiczne zaniedbania czy inne formy anosognozji jest taka, że niektóre marudy zostały u nich zabite. Tę teorię zaproponował, nie tak obrazowo, neuropsycholog Marcel Kinsbourne (1980), który wewnętrzne marudy nazwał „analizatorami korowymi". Z punktu widzenia modelu, który opracowujemy, zaniedbanie mogłoby być opisane jako utrata politycznego znaczenia przez pewne partie demonów w mózgu, spowodowana w wielu, ale nie wszystkich przypadkach śmiercią lub stłumieniem ich reprezentanta. Te demony nadal są aktywne, próbując wykonywać swoje czynności, a nawet czasami osiągając sukces, lecz nie mogą już wygrać w pewnych konkurencjach z lepiej zorganizowanymi koalicjami.

W tym modelu nieszkodliwe zaniedbanie naszych plamek ślepych niemal niedostrzegalnie przechodzi w przeróżne umiarkowanie dysfunkcyjne zaniedbania, na które cierpimy wszyscy, ale również w najbardziej zadziwiające zaniedbania, badane przez neurologów. Na przykład ja sam cierpię na kilka dosyć powszechnych rodzajów zaniedbywania. Najmniej poważne, choć czasem wstydliwe, jest moje *zaniedbanie literówek*. Jestem patologicznie niezdolny do dostrzegania tego rodzaju błędów w moich artykułach, gdy je sczytuję, i jedynie najbardziej pracowite ćwiczenia koncentracji sprawiają, że mogę to przezwyciężyć. Mój mózg, jak zasugerował Baars, nie „wypełnia" odpowiednich liter; nie musi „wypełniać", gdyż zwykle nie zwraca wystarczającej uwagi, aby zauważyć błędy; jego uwaga zostaje skierowana na inne właściwości słów na stronie. Kolejnym z moich małych upośledzeń jest *zaniedbanie oceniania studentów*. To niesamowite, jak atrakcyjna staje się perspektywa umycia podłogi w kuchni, zmiany ułożenia książek na półkach czy uporządkowania rachunków, gdy mam na biurku stertę testów egzaminacyjnych wymagających sprawdzenia. Ta właściwość, zwiększone zainteresowanie innymi możliwościami, jest szczególnie widoczna w zaniedbaniu jednostronnym; na pierwszy rzut oka, im dalej na prawo coś się znajduje, tym bardziej warte uwagi wydaje się pacjentowi z zaniedbywaniem lewostronnym. Być może jednak najbardziej poważny rodzaj zaniedbania, na który cierpię, to mój ciężki przypadek *zaniedbania finansów*.

Tak bardzo nie lubię wyliczać salda z książeczki czekowej, że jedynie jakaś naprawdę straszna perspektywa, na przykład ocenianie testów egzaminacyjnych studentów, może zmusić moją uwagę do skoncentrowania się na tym temacie. To zaniedbanie ma poważne konsekwencje dla mojego dobrobytu, konsekwencje, których istnienia jestem bardzo dobrze świadom, lecz pomimo tego zatrważająco nieskutecznego odwołania do mojej wewnętrznej racjonalności udaje mi się trwać w zaniedbaniu, chyba że zastosuję dosyć drastyczne sposoby automanipulacji.

Nie chodzi o to, że *nie widzę* książeczki czekowej, ale o to, że *nie chcę się jej przyglądać*. I choć w chłodnych, refleksyjnych momentach, takich jak ten, mogę o tym wszystkim opowiedzieć (co dowodzi, że nie mam *głębokiej* anosognozji związanej z tą niepełnosprawnością), zwykle po prostu nie dostrzegam własnego zaniedbania finansów. Krótko mówiąc, umiarkowana anosognozja. Z tej perspektywy jedyną zaskakującą rzeczą związaną z dziwnymi formami zaniedbania, badanymi przez neuropsychologów, jest kwestia granic. Wyobraźmy sobie kogoś, kto zaniedbuje wszystko po stronie *lewej* (Bisiach i Luzzatti 1978; Bisiach 1988; Bisiach i Vallar 1988; Calvanio, Petrone i Levine 1987). Albo kogoś, kto stracił widzenie barwne, *ale się na to nie skarży* (Geschwind i Fusillo 1966). A nawet kogoś, kto stracił wzrok, ale nadal nie zauważył tej głębokiej straty – zespół Antona, czyli zaprzeczanie ślepocie (Anton 1899; McGlynn i Schacter 1989, s. 154–158).

Te zaburzenia są łatwe do wyjaśnienia przez teorię świadomości modelu wielokrotnych szkiców, gdyż centralny świadek został zastąpiony koalicjami specjalistów, których szczególne głody epistemologiczne nie mogą zostać natychmiast przejęte przez innych agentów, jeśli są oni zlikwidowani lub na wakacjach[116]. Gdy te głody epistemologiczne znikają, to znikają bez śladu, pozostawiając pole innym koalicjom, innym agentom z innymi programami działania.

Jednak ta sama zasada wyjaśniająca zaniedbanie daje konkurencyjne ujęcie „przegapiania *qualiów* wzrokowych" naszego wyobrażonego wirtuoza ślepowidzenia. Zasugerowałem, iż jest możliwe, że jeśli narzeka na nieobecności *qualiów*, to *być może* po prostu zauważa względny niedostatek informacji otrzymywanych obecnie od swojego wzroku i źle tę sytuację opisuje. Spekulowałem również, że gdybyśmy w jakiś sposób mogli zwiększyć „szybkość transmisji" zbieranych przez niego informacji, wówczas część różnicy, jeśli nie cała różnica, między jego rodzajem widzenia a normalnym widzeniem mogłaby zostać zniwelowana. Widzimy teraz, że kolejny, tańszy sposób zniwelowania tej różnicy to po prostu zmniejszenie głodu epistemologicznego lub w jakiś sposób uśmierzenie jego wizualnej ciekawości. W końcu, jeśli pacjent z zespołem Antona może być całkowicie niewidomy, ale jeszcze sobie tego nie uświadamia, odrobina zaniedbania w strategicznej lokalizacji mogłaby zmienić naszego pacjenta ze ślepowidzeniem, skarżącego się na utratę *qualiów* wzrokowych, w nieskarżącego się pacjenta, deklarującego, iż jego wzrok został przywrócony bez zarzutu. Mogłoby się wydawać, że wiedzielibyśmy lepiej, ale czy rzeczywiście? Czy czegokolwiek brakowałoby takiej osobie? Nie istnieje *wymysł* w normalnym widzeniu, więc nie może go brakować. Czego zatem szukamy?

7. Wirtualna obecność

Mamy poczucie realności, gdy każde zadane pytanie dotyczące naszych układów wzrokowych uzyskuje odpowiedź tak szybko, że wydaje się, jakby te odpowiedzi już tam były.

Marvin Minsky, 1985, s. 257

Po raz kolejny nieobecność reprezentacji nie jest tym samym co reprezentacja nieobecności. A reprezentacja obecności nie jest tym samym co obecność reprezentacji. Trudno w to jednak uwierzyć. Nasze przekonanie, że mamy jakąś *bezpośrednią znajomość* szczególnych

właściwości naszych przeżyć, jest jedną z najpotężniejszych intuicji, z którymi musi skonfrontować się każdy, kto próbuje stworzyć dobrą teorię świadomości. Osłabiam je, starając się podkopać jej autorytet, ale nadal pozostaje wiele do zrobienia. Otto podejmuje jeszcze jedną próbę:

Twoja uwaga na temat Marilyn na tapecie jest tak naprawdę wątpliwą obroną dualizmu. Bardzo przekonująco argumentujesz, że w mózgu nie ma setek Marilyn o wysokiej rozdzielczości, a następnie wyciągasz wniosek, że nigdzie nie ma żadnych Marilyn! Jednak ja uważam, że skoro *to, co ja widzę*, to setki Marilyn o wysokiej rozdzielczości, to *skoro*, jak twierdzisz, nie ma ich nigdzie w moim mózgu, muszą być gdzieś indziej – w moim niefizycznym umyśle!

Setki Marilyn na tapecie wydają się obecne w przeżyciu, wydają się znajdować w umyśle, nie tylko na ścianie. Ale skoro, jak wiemy, spojrzenie może zmienić się w ułamku sekundy, aby zbierać informacje z każdej części środowiska wizualnego, dlaczego mózg miałby w ogóle kłopotać się importowaniem tych wszystkich Marilyn? Dlaczego zwyczajnie nie pozwolić światu na ich przechowanie, za darmo, dopóki nie okażą się potrzebne?

Porównajmy mózg do biblioteki. Niektóre są gigantycznymi magazynami, zawierającymi w swoich wnętrzach miliony książek, a wszystkie są dosyć szybko dostępne z odpowiednich stosów i regałów. Inne biblioteki trzymają mniej książek pod ręką, ale posiadają szczodre i efektywne systemy dostępu, kupując każdą książkę, jakiej wymagają ich użytkownicy, lub pożyczając je z innych bibliotek, korzystając ze sprawnego międzybibliotecznego systemu wypożyczania. Jeśli nie przechowujesz książek w swojej siedzibie, opóźnienia w dostępie są większe, ale nie o wiele większe. Możemy sobie wyobrazić elektroniczny system wypożyczania międzybibliotecznego (korzystający z faksu lub plików komputerowych), który może otrzymać książkę ze świata zewnętrznego szybciej, niż najbardziej sprawny biegacz mógłby zabrać książkę z regałów. Informatyk mógłby powiedzieć o książkach w takim systemie, że są cały czas „wirtualnie obecne" w bibliotece lub że „wirtualna kolekcja" jest setki lub tysiące razy większa niż jego rzeczywista kolekcja druków.

Jak moglibyśmy my, jako użytkownicy naszych własnych mózgów-bibliotek, wiedzieć, które z wydostawanych stamtąd elementów były *tam* cały czas, a które nasz mózg pobrał szybko ze świata zewnętrznego? Uważne eksperymenty przeprowadzone zgodnie z metodą heterofenomenologiczną mogą odpowiedzieć na to pytanie, ale sama introspekcja niewiele nam powie. Nie zatrzymuje nas to jednak przed myśleniem, że wiemy. W przypadku takiej czy innej nieobecności jakiegokolwiek dowodu nasza naturalna tendencja to dochodzenie do wniosku, że obecne jest więcej. Nazwałem to „pułapką introspekcji" (Dennett 1969, s. 139–140), a Minsky nazywa to „iluzją immanencji": „W przypadku gdy potrafisz odpowiedzieć na pytanie bez zauważalnego opóźnienia, wydaje się, że ta odpowiedź była już wcześniej aktywna w twoim umyśle" (Minsky 1985, s. 155).

System wypożyczania międzybibliotecznego to użyteczna, choć niekompletna analogia, gdyż twój mózg po prostu nie posiada możliwości uzyskania informacji na każdy zewnętrzny temat, jaki cię interesuje; ma dosłownie miliony wartowników spoglądających na jakąś część zewnętrznego świata, gotowych do wszczęcia alarmu i *przykucia* twojej uwagi do wszelkich nowych i ważnych zdarzeń następujących w świecie. W przypadku wzroku dzieje się to dzięki znajdującym się na siatkówce, wokół centralnej części dołka środkowego, czopkom i pręcikom oraz nerwowym agentom ulokowanym na pokładzie tych wartowniczych wież, specjalizujących się w wykrywaniu zmian w ruchu. Jeśli jeden z tych agentów podniesie alarm – „Zmiana w moim sektorze!" – to niemal w tym samym momencie zostaje rozpoczęta sakkada, prowadząca dołek środkowy do skierowania się na odpowiedni region, i nowość można zlokalizować,

zidentyfikować i się nią zająć. System strażników jest tak niezawodny, że jest bardzo trudno przemycić zmianę do widzialnego świata bez informowania o tym całego układu wzrokowego, lecz z pomocą zaawansowanych technologicznie sztuczek wartownicy mogą czasem zostać ominięci, czego rezultaty są niezwykłe.

Gdy oczy błądzą w ruchach sakkadowych, skurcze mięśni powodujące, że gałki oczne się obracają, to czynności balistyczne: punkty fiksacji to pozbawione prowadzenia pociski, których trajektorie w momencie rozpoczęcia determinują, gdzie i kiedy dotrą do strefy zero, w której znajduje się nowy cel. Kiedy na przykład czytasz tekst na ekranie komputera, oczy przetoczą się po kilku słowach z każdą sakkadą, tym dalej i szybciej, im lepiej czytasz. Jak by to było, gdyby magik w rodzaju kartezjańskiego złego demona na skromną skalę mógł zmienić świat na kilka milisekund, podczas których twoje oczy skaczą do kolejnego celu? W sposób zdumiewający komputer wyposażony w automatyczny okulograf może wykryć i zanalizować start w pierwszych milisekundach sakkady, obliczyć, gdzie znajdzie się strefa zero oraz, *zanim sakkada się zakończy*, usunąć słowo z ekranu w strefie zero i zastąpić je innym słowem tej samej długości. Co zobaczysz? Tylko nowe słowo, i to bez żadnego poczucia, że cokolwiek się przydarzyło. Gdy przeglądasz tekst na ekranie, wydaje ci się z całą pewnością tak stabilny, jakby słowa były wykute w marmurze, ale innej osobie czytającej ten sam tekst przez twoje ramię (i inaczej kierującej swoimi ruchami sakkadowymi) ekran wydaje się drżeć od zmian.

Efekt jest oszałamiający. Kiedy po raz pierwszy zetknąłem się z eksperymentem z okulografem i zobaczyłem, jak nieświadomi (najwyraźniej) zmian zachodzących na ekranie byli badani, poprosiłem o to, by i mnie zbadano. Chciałem sam tego doświadczyć. Posadzono mnie przed komputerem, a moja głowa została unieruchomiona przez zaciśnięcie zębów na gryzaku. W ten sposób łatwiej jest pracować okulografowi, który odbija niezauważalny promień światła od soczewki oka badanego i analizuje dane w poszukiwaniu ruchu oka. Czekając na to, aby eksperymentatorzy włączyli urządzenie, czytałem tekst na ekranie. Czekałem i czekałem, bardzo chciałem już zacząć. Stałem się niecierpliwy. „Dlaczego go nie włączacie?" – zapytałem. „Jest włączony" – odpowiedzieli.

Wszystkie zmiany na ekranie zachodzą podczas sakkad, więc strażnicy nie są w stanie wszcząć odpowiedniego alarmu. Jeszcze niedawno zjawisko to znane było jako „tłumienie wzrokowe". Uważano, że mózg musi w jakiś sposób odciąć się od wszelkiej informacji przychodzącej z oczu podczas sakkad, ponieważ nie potrafimy zauważyć zmian zachodzących w polu widzenia, gdy one trwają, i oczywiście nikt nie skarży się na zawrotne i zatrważające zmiany. Jednak sprytny eksperyment z okulografem (Brooks i in. 1980) pokazał, że jeśli bodziec – taki jak słowo czy litera alfabetu – zostaje ruszony jednocześnie z sakkadą, trzymając tempo „cienia" dołka środkowego, gdy pędzi on na nowe miejsce lądowania, jest bez problemu widziany i identyfikowany przez badanego. Dane przychodzące z oka nie są blokowane w drodze do mózgu podczas sakkad, ale w normalnych warunkach są one nie do użytku – po prostu wszystko pędzi zbyt szybko, aby coś z tego wyciągnąć – więc mózg nieszkodliwie to zaniedbuje. Jeśli wszyscy wartownicy wysyłają alarm w tym samym czasie, najlepsze, co można zrobić, to po prostu ich zignorować.

W sytuacji eksperymentalnej, w której się znalazłem, słowa na ekranie były wymazywane i zastępowane podczas moich sakkad. Jeśli twoje widzenie na obrzeżach dołka środkowego nie potrafi zidentyfikować słowa w strefie zero, zanim rozpocznie się ruch sakkadowy w jej kierunku, w momencie gdy już dotrzesz i je zidentyfikujesz, nie może istnieć żaden wcześniejszy jego zapis czy wspomnienie, z którym mógłbyś je porównać. Zmiana nie może zostać dostrzeżona, ponieważ informacja logicznie wymagana do takiej obserwacji po prostu *nie istnieje*. Oczywiście *wydaje ci się*, gdy czytasz tę stronę, że wszystkie słowa w linii są w pewien

sposób obecne w twojej świadomości (na drugim planie), nawet zanim zajmiesz się konkretnie nimi, jednak jest to iluzja. Są obecne jedynie wirtualnie.

Istnieją oczywiście *pewne* dane w mózgu dotyczące słów znajdujących się wokoło – wystarczające na przykład, aby pokierować i wywołać najbliższą sakkadę. Ale jakie dokładnie informacje już *tam* są? Eksperymenty z okulografami oraz podobnymi urządzeniami mogą wyznaczyć granice tego, co potrafisz dostrzec, i tym samym wyznaczyć granice tego, co jest *obecne* w umyśle. (Zob. np. Pollatsek, Rayner i Collins 1984; Morris, Rayner i Pollatsek 1990). Twierdzenie, które kusiło Ottona, że to, co nie znajduje się w mózgu, musi mimo wszystko znajdować się w umyśle, gdyż z pewnością *wydaje się* tam być, jest jałowe. A to dlatego, że, jak widzieliśmy, nie znajdowałoby się „tam" w żadnym sensie, który mógłby wpłynąć na *własne przeżycia Ottona*, a już na pewno nie na jego możliwości zdania testu, wciskania przycisków itd.

8. Zobaczyć to uwierzyć: dialog z Ottonem

W tym momencie nasz krytyk Otto domaga się podsumowania, gdyż jest pewien, że został gdzieś po drodze oszukany. Zamierzam przeprowadzić z nim dialog, mając nadzieję, że rozwieję również wiele, jeśli nie wszystkie wasze wątpliwości. Zaczyna Otto:

Wydaje mi się, że zaprzeczyłeś istnieniu niewątpliwie najbardziej prawdziwych zjawisk, jakie znamy: rzeczywistego wydawania się, w które nie mógł zwątpić nawet Kartezjusz w swoich *Medytacjach*.

W pewnym sensie masz rację; przeczę właśnie jego istnieniu. Wróćmy do zjawiska neonowej poświaty. Wydaje się, że na tylnej okładce dostrzegamy różowawy, żarzący się pierścień.

Rzeczywiście tak jest.

Ale tak naprawdę nie ma żadnego różowawego, żarzącego się pierścienia.

Racja. Ale z pewnością wydaje się, że jest!

To prawda.

A więc gdzie on jest?

Gdzie jest co?

Różowawy, żarzący się pierścień.

Nie ma go; wydawało mi się, że właśnie to przyznałem.

Cóż, owszem, nie ma żadnego różowawego, żarzącego się pierścienia na tej stronie, ale przecież wydaje się, że jest.

Racja. Wydaje się, że jest tam różowawy, żarzący się pierścień.

Porozmawiajmy więc o *tym* pierścieniu.

O którym?

O tym, który *wydaje* się, że jest.

Nie ma czegoś takiego jak różowy pierścień, który tylko wydaje się, że jest.

Posłuchaj, nie tylko *mówię*, że tam wydaje się być różowawy, żarzący się pierścień: *on tam rzeczywiście wydaje się być*!

Oczywiście się z tobą zgadzam. Nigdy nie oskarżyłbym cię o nieszczerość! Naprawdę masz na myśli to, że tam wydaje się być różowawy, żarzący się pierścień.

Posłuchaj, ja nie tylko mam to na myśli. Nie tylko *uważam*, że wydaje się tam być różowawy, żarzący się pierścień; *tam rzeczywiście* wydaje się istnieć różowawy, żarzący się pierścień!

Zrobiłeś to. Wpadłeś w pułapkę, wraz z wieloma innymi. Wydaje się, że myślisz, że jest różnica między uważaniem (ocenieniem, decydowaniem, byciem przekonanym), że coś wydaje

się tobie różowe, a tym, że coś *naprawdę wydaje się tobie* różowe. Ale nie ma tu różnicy. Nie ma takiego zjawiska jak prawdziwe wydawanie się – obok zjawiska oceniania w taki czy inny sposób, że coś jakieś jest.

Przypomnij sobie tapetę z Marilyn. Ściana rzeczywiście jest pokryta obrazami Marilyn w wysokiej rozdzielczości. Co więcej, właśnie tak ci się wydaje! Wydaje ci się, że ściana jest pokryta obrazami Marilyn w wysokiej rozdzielczości. Masz szczęście, aparat wzrokowy doprowadził cię do prawidłowego przekonania o właściwościach twojego środowiska. Jednak nie ma mnóstwa *prawdziwie wydających się* obrazów Marilyn reprezentowanych w twoim mózgu – lub umyśle. Nie ma nośnika, który *odbija* szczegóły tapety, który *przedstawia* ją dla twojego wewnętrznego świadka. Po prostu wydaje ci się, że jest tam mnóstwo obrazów Marilyn w wysokiej rozdzielczości (i w tym przypadku masz rację – rzeczywiście tam są). W innych przypadkach możesz się mylić; może ci się wydawać – w zjawisku kolorowego phi – że pojedynczy punkt przesunął się w prawo, zmieniając po drodze kolor, gdy w rzeczywistości były to po prostu dwa błyskające punkty w różnych kolorach. To, że tak ci się wydaje, nie wymaga *przedstawiania* w mózgu bardziej niż to, że ocena barwy przez mózg, gdy już zostanie uzyskana, nie musi być następnie gdzieś *odkodowana*.

Ale w takim razie co się dzieje, gdy wydaje mi się, że jest tam różowawy, żarzący się pierścień? Jakie pozytywne ujęcie przedstawia twoja teoria? Wydaje mi się, że w tej kwestii jesteś bardzo wymijający.

Chyba masz rację. Czas, żeby wszystko wyśpiewać i zaprezentować pozytywne ujęcie, jednak wyznaję, że będę musiał to zrobić, zaczynając od karykatury, następnie ją dopiero rewidując. Nie potrafię chyba znaleźć bardziej bezpośredniego sposobu na wyjaśnienie tego wszystkiego.

Zauważyłem. Kontynuuj.

Załóżmy, że *istnieje* centralny nadawacz sensów. Załóżmy jednak, że zamiast siedzieć w teatrze kartezjańskim i oglądać przedstawienia, centralny nadawacz sensów siedzi w ciemności i ma przeczucia – właśnie ni z tego, ni z owego stwierdził, że jest tam coś różowego, w taki sposób, w jaki mogłoby ci się nagle wydać, że ktoś za tobą stoi.

Jakie to są dokładnie przeczucia? Z czego są zrobione?

Dobre pytanie, na które początkowo muszę odpowiedzieć wymijająco, poprzez karykaturę. Te przeczucia to sądy, które centralny nadawacz sensów wykrzykuje do siebie w swoim własnym, specjalnym języku myśleńskim. Zatem jego życie składa się z serii osądów, które są zdaniami w myśleńskim wyrażającymi kolejno różne sądy z ogromną prędkością. Niektóre z nich decyduje się publikować w angielskim tłumaczeniu.

Ta teoria ma zaletę polegającą na pozbyciu się *wymysłu*, rzutowania w przestrzeń fenomenalną, wypełnienia luk na ekranie teatru, nadal jednak ma centralnego nadawacza sensów oraz myśleński. Zrewidujmy więc teraz tę teorię. Najpierw pozbądźmy się centralnego nadawacza sensów, rozpraszając wszystkie jego oceny w czasie i przestrzeni w mózgu – każdy akt rozróżnienia lub ustalenia treści gdzieś się zdarza, ale nie ma *jednego* rozróżniacza, który wykonuje całą tę pracę. Następnie pozbądźmy się myśleńskiego; treść ocen nie musi być wyrażalna w formie „sądów" – to błąd, przypadek zbyt entuzjastycznego i nieprawidłowego rzutowania kategorii językowej z powrotem na czynności mózgowe.

Zatem przeczucia są jak akty mowy z wyjątkiem tego, że nie ma aktora i mowy!

No cóż, tak. To, co rzeczywiście jest, to różne zdarzenia ustalenia treści odbywające się w różnych miejscach i w różnych momentach w mózgu. Nie są to niczyje akty mowy, więc nie muszą odbywać się w języku, jednak są raczej *jak* akty mowy; mają treść oraz efekt informowania różnych procesów o swojej treści. Rozważaliśmy bardziej szczegółowe przypadki

w rozdziałach 5–10. Niektóre z tych ustaleń treści mają dalsze efekty, które w końcu prowadzą do wypowiedzenia zdań – w języku naturalnym – publicznego lub jedynie wewnętrznego. I w taki sposób zostaje stworzony tekst heterofenomenologiczny. Gdy zostaje zinterpretowany, wykreowana zostaje korzystna iluzja autora. Wystarczy to do stworzenia heterofenomenologii.

Ale co z *prawdziwą* fenomenologią?

Nic takiego nie istnieje. Przypomnij sobie nasze rozważania o interpretacji fikcji. Gdy napotykamy na powieść, która jest delikatnie zawoalowaną autobiografią, stwierdzamy, że jesteśmy w stanie połączyć fikcyjne wydarzenia z wieloma prawdziwymi wydarzeniami z życia autora, więc w naciągniętym sensie powieść ta jest o tych prawdziwych wydarzeniach. Autor może w ogóle nie zdawać sobie z tego sprawy, jednak mimo wszystko w tym naciąganym sensie to prawda; to tych wydarzeń dotyczy tekst, gdyż są to prawdziwe wydarzenia, wyjaśniające, dlaczego *ten* tekst został stworzony.

Ale o czym jest tekst w sensie *nie*naciąganym?

O niczym. To fikcja. *Wydaje się* o różnych fikcyjnych postaciach, miejscach i wydarzeniach, ale nigdy się one nie wydarzyły; tak *naprawdę* jest o niczym.

Kiedy jednak czytam powieść, to te postacie fikcyjne stają się żywe! Coś się ze mną dzieje; *wizualizuję* wydarzenia. Akt czytania oraz interpretacji tekstu takiego jak powieść *tworzy* coś nowego w mojej wyobraźni: obrazy postaci dokonujących jakichś czynów. No bo gdy idziemy zobaczyć filmową wersję powieści, którą przeczytaliśmy, często myślimy: „Zupełnie inaczej sobie ją wyobrażałem!".

Masz rację. W *Fearing Fictions* filozof Kendall Walton twierdzi, że te akty wyobraźni ze strony interpretującego dopełniają tekst w takim samym sensie jak obrazki, które można znaleźć w ilustrowanych wydaniach powieści, „łącząc się z powieścią, tworzą »większy« [fikcyjny, heterofenomenologiczny] świat" (Walton 1978, s. 17). Te dodatki są zupełnie prawdziwe, ale są po prostu dalszym „tekstem" – nie są zrobione z wymysłu, lecz z oceny. Fenomenologia to nic ponadto.

Ale wydaje się inaczej!

Właśnie! *Wydaje się, że istnieje fenomenologia*. Jest to fakt entuzjastycznie przyznawany przez heterofenomenologa. Jednak *nie* wynika z tego niezaprzeczalnego, uniwersalnie stwierdzanego faktu, że *rzeczywiście istnieje* fenomenologia. W tym tkwi sedno sprawy.

Czy zatem przeczysz temu, że świadomość *jest wypełniona*?

Owszem. To część tego, co neguję. Świadomość jest pełna luk i rozrzedzona i nie zawiera połowy rzeczy, które ludzie myślą, że zawiera.

Ale, ale...

Ale przecież świadomość wydaje się wypełniona?

Tak!

Zgadzam się; wydaje się wypełniona; wydaje się nawet, że „uderzający fakt" dotyczący świadomości to właśnie to, że jest ciągła, jak mówi Edelman, ale...

Wiem, wiem: z faktu, że *wydaje się* wypełniona, nie wynika to, że *jest* wypełniona.

Teraz już rozumiesz.

Ale mam jeszcze jeden problem z tym przejściem pełnym luster, które nazywasz teorią. Mówisz, że jest jedynie *tak, jakby* istniał centralny nadawacz sensów, *tak jakby* istniał pojedynczy autor, *tak jakby* było miejsce, w którym to wszystko się ze sobą łączy! Nie rozumiem całej tej kwestii *tak jakby*!

Być może kolejny eksperyment myślowy pomoże ci to pojąć. Wyobraź sobie, że odwiedziliśmy inną planetę i dowiedzieliśmy się, że tamtejsi naukowcy mają dosyć uroczą teorię: każda fizyczna rzecz ma wewnątrz duszę, a każda dusza kocha wszystkie inne dusze. W związku

z tym rzeczy zwykle poruszają się w swoim kierunku, nakłaniane do tego przez miłość swoich wewnętrznych dusz do siebie. Co więcej, możemy sobie wyobrazić, że naukowcy wypracowali dosyć udany system lokalizowania duszy i w momencie gdy stwierdzają dokładnie, gdzie się ona znajduje w przestrzeni fizycznej, są w stanie odpowiedzieć na pytania dotyczące stabilności istoty („Przewróci się, bo jego dusza znajduje się tak wysoko"), wibracji („Jeśli ułożysz przeciwważący przedmiot po drugiej stronie tego koła, które ma dość sporą duszę, zminimalizuje on chwianie") oraz wielu innych kwestii technicznych.

Oczywiście moglibyśmy im powiedzieć, że odkryli pojęcie środka ciężkości (lub, dokładniej mówiąc, środka masy) i że traktują je trochę zbyt ceremonialnie. Mówimy im, że mogą nadal mówić i myśleć jak dotychczas – muszą jedynie pozbyć się odrobiny niepotrzebnego bagażu metafizycznego. Istnieje prostsza, oszczędniejsza (i o wiele bardziej satysfakcjonująca) interpretacja faktów, których używają do zrozumienia swojej fizyki duchowej. Pytają nas: Czy to są dusze? Cóż, tak – odpowiadamy – ale są *abstrakcyjne*, są abstraktami matematycznymi, a nie okruchami tajemniczej substancji. To niezwykle użyteczne fikcje. Jest *tak, jakby* każdy obiekt przyciągał inny, koncentrując swoją grawitacyjną moc w konkretnym punkcie – i jest nieporównanie łatwiej przewidzieć zachowanie systemów, korzystając z tej wynikającej z zasad fikcji niż przy użyciu niewygodnych szczegółów – każdy punkt przyciągający każdy inny punkt.

Czuję się *tak, jakbym* właśnie został oszukany.

Nie mów, że cię nie ostrzegałem. Nie możesz się spodziewać, że świadomość okaże się *dokładnie* tym, czego się spodziewałeś. Poza tym czego tak naprawdę się pozbywasz?

Tylko mojej duszy.

Nie w żadnym spójnym, dającym się obronić sensie. Pozbywasz się tylko odrobiny czegoś szczególnego, co i tak nie mogło być tak naprawdę szczególne. Dlaczego miałbyś myśleć o sobie lepiej, gdyby się okazało, że jesteś swego rodzaju umysłową perłą w muszli mózgu? Cóż takiego specjalnego byłoby w umysłowej perle?

Umysłowa perła mogłaby być nieśmiertelna, w przeciwieństwie do mózgu.

Idea, że jaźń – czy dusza – jest jedynie abstrakcją, przez wielu odbierana jest jako idea negatywna, jako zaprzeczenie, a nie coś pozytywnego. Ale tak naprawdę ma w sobie wiele atutów, łącznie z – jeśli jest to dla ciebie istotne – poniekąd zdrowiej pojmowaną wersją potencjalnej nieśmiertelności, niż cokolwiek, co można znaleźć w tradycyjnych ideach duszy, jednak będzie to musiało poczekać do rozdziału 13. Najpierw w sposób definitywny musimy poradzić sobie z *qualiami*, które *nadal* są zakorzenione w naszej wyobraźni.

Rozdział 12

Dyskwalifikacja *qualiów*

1. Nowy sznurek do latawca

Wrzucone w lukę przyczynową,
quale po prostu przez nią przeleci.
Ivan Fox, 1989, s. 82

Gdy sznurek przy twoim latawcu się poplącze, w zasadzie można go odplątać, zwłaszcza jeśli masz cierpliwość i analityczny umysł. Istnieje jednak granica, po której przekroczeniu zasada przestaje działać i triumfuje praktyczność. Niektóre plątaniny trzeba wyrzucić. Poszukajmy nowego sznurka do latawca. Tak naprawdę okazuje się to tańsze niż praca, jaką musielibyśmy włożyć w ratowanie starego sznurka, a poza tym latawiec szybciej może znów latać. Tak właśnie jest, według mnie, z filozoficznym tematem *qualiów*, męczącym splotem coraz bardziej zawiłych i przedziwnych eksperymentów myślowych, żargonu, dowcipów, aluzji do domniemanego obalenia zarzutów, „tradycyjnych" wyników, które nie powinny należeć do żadnej tradycji, i mnóstwa innych zbijających z tropu i marnujących czas bzdur. Czasem od bałaganu najlepiej jest odejść, więc nie zamierzam przeprowadzać analitycznej kwerendy literatury przedmiotu, chociaż zawiera momenty głębokiego namysłu i pomysłowości, które dały mi wiele pożytku (Shoemaker 1975, 1981, 1988; White 1986; Kitcher 1979; Harman 1990; Fox 1989). Próbowałem już w przeszłości rozplątać tę kwestię (Dennett 1988a), ale teraz myślę, że będzie lepiej, jeśli zaczniemy od samego początku.

Nietrudno zauważyć, dlaczego filozofowie tak się plątają w kwestii *qualiów*. Zaczęli od tego, od czego zaczęłaby każda rozumna osoba: od swoich najsilniejszych i najwyraźniejszych intuicji dotyczących ich własnych umysłów. Niestety, te intuicje tworzą wzajemnie się wspierający, zamknięty krąg doktryn i zamykają wyobraźnię w teatrze kartezjańskim. Mimo że filozofowie odkryli paradoksy nieodłącznie związane z tym zamkniętym kręgiem idei – dlatego istnieje literatura o *qualiach* – nie mają *pełnej konkurencyjnej wizji*, którą mogliby przyjąć, więc opierając się na swoich nadal silnych intuicjach, dali się z powrotem wciągnąć do więzienia paradoksów. Właśnie dlatego literatura dotycząca *qualiów* robi się coraz bardziej zagmatwana, zamiast dojść do jakiegoś konsensusu. Ale teraz przedstawiliśmy właśnie taką alternatywną wizję, model wielokrotnych szkiców. Korzystając z niego, możemy zaproponować trochę inne, pozytywne ujęcie tej kwestii. Następnie możemy zrobić przerwę w sekcjach 4 i 5, aby porównać ją do tych wizji, które, mam nadzieję, zostaną przez nią zastąpione.

Pewna świetna książka wprowadzająca w zagadnienia neurobiologii zawiera następujący ustęp:

„Barwa" jako taka nie istnieje w świecie; istnieje jedynie w oku i mózgu obserwatora. Przedmioty odbijają wiele różnych długości fal świetlnych, ale te fale same w sobie nie mają barwy. [Ornstein i Thompson 1984, s. 55]

Jest to trafny cios w zdrowy rozsądek, jednak zauważmy, że gdyby wziąć go ściśle i dosłownie, nie mógłby być tym, co autorzy mają na myśli, nie mógłby być prawdą. Mówią, że barwa nie istnieje „w świecie", tylko w „oczach i mózgu" obserwatora. Ale oczy i mózg

obserwatora należą do świata, są jego częścią dokładnie tak samo jak widziane przez niego obiekty. I tak jak te obiekty, oczy i mózg mają barwy. Oczy mogą być niebieskie, brązowe czy zielone, a nawet mózg nie jest zrobiony *jedynie* z substancji szarej (i białej): poza *substantia nigra* (istotą czarną) jest w nim *locus coeruleus* (miejsce sinawe). Ale oczywiście barwy, które są w „oku i mózgu obserwatora", w *tym* sensie nie są tym, co autorzy mieli na myśli. Co sprawia, że ktoś może pomyśleć, iż istnieje barwa w jakimkolwiek innym sensie?

Nowoczesna nauka – tak zwykle się powiada – usunęła barwę ze świata fizycznego, zastępując ją promieniowaniem elektromagnetycznym o różnej długości fal docierających do powierzchni, które w charakterystyczny dla siebie sposób je odbijają i wchłaniają. Może się wydawać, że barwa *tam* jest, ale jej tam nie ma. Jest *tutaj* – w „oczach i mózgu obserwatora". (Gdyby autorzy tego ustępu nie byli tak dobrymi materialistami, prawdopodobnie powiedzieliby, że jest ona w *umyśle* obserwatora, ratując się przed głupkowatą interpretacją, którą właśnie odrzuciliśmy, ale stwarzając jeszcze gorsze problemy dla siebie). Skoro jednak nie ma wewnętrznego *wymysłu*, który mógłby zostać zabarwiony w jakimś specjalnym, subiektywnym, wewnątrzumysłowym, fenomenalnym sensie, wydaje się, że barwy całkowicie znikają! *Coś* musi być barwami, które znamy i kochamy, barwami, które mieszamy i dopasowujemy. Gdzie, ach, gdzie może to być?

Jest to stary, filozoficzny dylemat, z którym musimy się teraz zmierzyć. W XVII wieku filozof John Locke (a przed nim naukowiec Robert Boyle) nazwał takie cechy jak kolory, zapachy, smaki i dźwięki „*własnościami wtórnymi*". Były one różne od własności pierwotnych: wielkości, kształtu, ruchu, ilości i twardości. Cechy wtórne nie były rzeczami w umyśle, lecz raczej *siłami rzeczy* w świecie (dzięki ich szczególnym właściwościom pierwotnym), powodującymi czy wywołującymi pewne rzeczy w umysłach zwykłych obserwatorów. (A co, jeśli nie byłoby w pobliżu żadnego obserwatora? Jest to odwieczna zagadka o drzewie w lesie, które się przewraca. Czy wydaje jakiś dźwięk? Odpowiedź pozostawiam czytelnikowi jako ćwiczenie). Locke'owska definicja cech wtórnych stała się częścią standardowej interpretacji nauki przez laika i ma swoje zalety, ale również cenę: rzeczy wytworzone w umyśle. Na przykład cecha wtórna *czerwień* była dla Locke'a dyspozycyjną cechą lub mocą pewnych powierzchni obiektów fizycznych, dzięki ich cechom mikrostrukturalnym, do wytworzenia w nas *idei czerwieni* wtedy, gdy światło jest odbite od tych powierzchni do naszych oczu. Moc zewnętrznego przedmiotu wydaje się dosyć jasna, ale czym jest idea czerwieni? Czy jest jak piękna niebieska suknia, *zabarwiona* – w jakimś sensie? Czy też jest, jak piękne rozważania o fiolecie, jedynie o barwie, a sama barwy w ogóle nie posiada? Otwiera to pewne możliwości, ale w jaki sposób idea może być tylko *o* barwie (na przykład o czerwieni), jeśli nic nigdzie nie *jest* czerwone?

Co to właściwie jest czerwień? Czym są barwy? Barwy zawsze były ulubionymi przykładami filozofów i na razie będę trzymał się tej tradycji. Główny jej problem elegancko wyłania się z filozoficznej analizy Wilfrida Sellarsa (1963/1991, 1981b), który odróżniał dyspozycyjne cechy obiektów (cechy wtórne Locke'a) od tego, co nazywał „własnościami bieżącymi". Różowa kostka lodu w zamrażalniku ze zgaszonym światłem posiada cechę wtórną różowości, jednak nie ma przykładu własności *bieżącej różowości*, dopóki obserwator nie otworzy drzwi i nie spojrzy na kostkę. Czy bieżący różowy jest właściwością czegoś w mózgu, czy czegoś „w świecie zewnętrznym"? Sellars twierdził, że w każdym z przypadków bieżący różowy jest „homogeniczną" właściwością czegoś rzeczywistego. Twierdzeniem o homogeniczności negował między innymi hipotezę, że bieżący różowy jest czymś w rodzaju *neuronalnej aktywności o natężeniu 97 w regionie 75* w mózgu. Chciał również zaprzeczyć temu, że subiektywny świat fenomenologii barw wyczerpuje się w czymś tak pozbawionym kolorów,

jak *osądy* dotyczące tego, że taka czy inna rzecz jest lub wydaje się różowa. Na przykład akt przypominania sobie oczyma wyobraźni barwy dojrzałego banana i osąd, że jest to kolor żółty, same w sobie nie przywołałyby istnienia bieżącej barwy żółtej (Sellars 1981; Dennett 1981b). Byłyby jedynie osądzaniem, że coś jest żółte, zjawiskiem samym w sobie pozbawionym bieżącego żółtego, tak jak w przypadku wiersza o bananach.

Sellars uważał nawet, iż wszystkie nauki fizyczne musiałyby zostać zrewolucjonizowane, aby zrobić miejsce bieżącemu różowi i jego krewnym. Niewielu filozofów poparło go w tym radykalnym poglądzie, ale pewna jego wersja została ostatnio wskrzeszona przez filozofa Michaela Lockwooda (1989). Inni filozofowie, jak Thomas Nagel, założyli, że nawet zrewolucjonizowana nauka nie byłaby w stanie poradzić sobie z takimi właściwościami:

> Subiektywne cechy świadomych procesów psychicznych – w odróżnieniu od ich fizycznych przyczyn i skutków – wymykają się oczyszczonej formie myślenia, radzącej sobie ze światem fizycznym, leżącym u podłoża zjawisk. [Nagel 1986/1997, s. 21–22]

Filozofowie przyjęli różne nazwy na rzeczy w obserwatorze (lub na cechy obserwatora), które miały zapewnić bezpieczną przystań barwom i reszcie właściwości wygnanych z „zewnętrznego świata" przez triumfy fizyki: „surowym czuciom", „doznaniom", „własnościom fenomenalnym", „własnościom wewnętrznym przeżyć świadomych", „jakościowej treści stanów umysłowych" oraz, oczywiście, „*qualiom*", pojęciu, którego będę używał. Między definicjami tych pojęć istnieją subtelne różnice, lecz zamierzam jeździć po nich jak po łysej kobyle. Wydawało się, że w poprzednim rozdziale zaprzeczyłem istnieniu *jakichkolwiek* tego rodzaju właściwości, i w tym przypadku to, co się wydaje, *takie* właśnie jest. *Zaprzeczam* istnieniu takich właściwości. Ale (oto znów powraca ten sam motyw) całym sercem zgadzam się z tym, że *qualia* wydają się istnieć.

Qualia wydają się istnieć, ponieważ rzeczywiście zdaje się, że nauka pokazała nam, iż barwy nie mogą istnieć tam, a zatem muszą być tu. Co więcej, wydaje się, że to, co jest tutaj, nie może *po prostu* być osądem, jakiego dokonujemy, gdy przedmioty wydają się nam barwne. To, co rzeczywiście pokazała nam nauka, to jedynie fakt, że właściwości przedmiotów związane z odbijaniem światła sprawiają, iż istoty żywe wchodzą w różne stany rozróżniania, rozproszone w ich mózgach, będące podstawą mnóstwa wewnętrznych dyspozycji i wyuczonych nawyków o różnej złożoności. A jakie są *ich* właściwości? Tu możemy zagrać kartą Locke'a po raz drugi: te charakterystyczne stany mózgów obserwatorów mają różne własności „pierwotne" (cechy mechaniczne związane z ich połączeniami, stany pobudzenia ich elementów itp.) i ze względu na te własności pierwotne mają cechy wtórne, które są jedynie dyspozycyjne. Na przykład u istot ludzkich posługujących się językiem te charakterystyczne stany często w końcu nakłaniają istoty do wyrażenia werbalnej oceny nawiązującej do „barwy" różnych obiektów. Gdy ktoś mówi: „Wiem, że pierścień nie jest tak naprawdę różowy, ale z pewnością taki się wydaje", pierwsza część tej wypowiedzi wyraża sąd o czymś w świecie, a druga wyraża sąd drugiego rzędu dotyczący charakterystycznego stanu czegoś w świecie. Semantyka takich stwierdzeń pokazuje, czym rzekomo są barwy: właściwościami odbijania światła od powierzchni przedmiotów lub od przezroczystych obiektów (różowa kostka lodu, snop światła). I właśnie tym w rzeczywistości są – choć stwierdzenie, *które* dokładnie właściwości odbijania mają, jest trudne (z powodów, o których powiemy sobie w kolejnym podrozdziale).

Czy nasze wewnętrzne stany charakterystyczne *również* mają pewne specjalne „wewnętrzne" własności, subiektywne, prywatne, niewyrażalne, które stanowią o tym, *jakie wydają nam się rzeczy* (jak dla nas brzmią, jak smakują itd.)? Te dodatkowe własności byłyby *qualiami*, ale zanim przyjrzymy się argumentom wysuniętym przez filozofów w celu *dowiedzenia*, że te dodatkowe własności istnieją, spróbujemy usunąć motywację wiary w ich

istnienie, szukając konkurencyjnych wyjaśnień zjawiska, które wydaje się ich wymagać. Wówczas systematyczne błędy w próbach dowodzenia będą wyraźnie widoczne.

Według poglądu konkurencyjnego barwy mimo wszystko *są* własnościami znajdującymi się „na zewnątrz". Zamiast „idei czerwoności" Locke'a mamy (u normalnych istot ludzkich) stany rozróżniania, które posiadają tę treść: *czerwień*. Przykład pomoże jasno zrozumieć, czym są te stany rozróżniania – a, co ważniejsze, czym nie są. Możemy porównać barwy przedmiotów w świecie, kładąc je obok siebie i patrząc na nie, aby wyrobić sobie jakiś osąd, możemy jednak także porównać barwy przedmiotów jedynie przez przypomnienie ich sobie lub przez ich wyobrażenie „w naszych umysłach". Czy standardowa barwa czerwonych pasków na fladze amerykańskiej to ten sam czerwony, czy może ciemniejszy, jaśniejszy, bardziej jaskrawy, bądź też mniej lub bardziej pomarańczowy niż standardowy czerwony płaszcza świętego Mikołaja (albo brytyjskiej skrzynki pocztowej lub radzieckiej czerwonej gwiazdy)? (Jeśli nie pamiętasz żadnej z par tych standardów, spróbuj innej pary, na przykład niebieskiego z karty Visa i niebieskiego nieba, zielonego z filcu na stole bilardowym i zielonego jabłka Granny Smith, czy też koloru żółtego cytryny i masła). Możemy dokonywać takich porównań „oczyma naszej wyobraźni", a gdy to robimy, to jakoś sprawiamy, że dzieje się w nas coś, co pobiera informacje z pamięci i pozwala nam porównać, w świadomym przeżyciu, barwy typowych przedmiotów wedle naszej pamięci (a w każdym razie takie, jak nam się wydaje, że je pamiętamy). Niektórzy z nas są w tym bez wątpienia lepsi niż inni, a wiele osób nie ma pewności co do osądów, jakich dokonujemy w takich przypadkach. Właśnie dlatego zabieramy do domu próbki farb lub bierzemy próbki materiałów do sklepu z farbami, aby móc położyć obok siebie, w świecie zewnętrznym, przykłady dwóch barw, które chcemy ze sobą porównać.

Gdy już dokonamy tych porównań „oczyma naszej wyobraźni", co się dzieje, według mojego poglądu? Coś ściśle analogicznego do tego, co wydarzyłoby się w maszynie – w robocie – która również mogłaby dokonywać takich porównań. Przypomnijmy sobie Vorsetzer systemu CAD Dla Niewidomych 1.0 z rozdziału 10 (tego z kamerą, która mogła być wycelowana w ekran systemu CAD). Załóżmy, że kładziemy przed nim barwny obrazek świętego Mikołaja i pytamy go, czy czerwony na obrazku jest głębszy niż czerwony z amerykańskiej flagi (coś, co już przechowuje w swojej pamięci). Oto co by zrobił: pobrałby reprezentację flagi z pamięci i zlokalizował na niej czerwone paski (są oznakowane jako „czerwony nr 163" na jego diagramie). Następnie porównałby tę czerwień z czerwienią płaszcza świętego Mikołaja z obrazka znajdującego się przed kamerą, który okazuje się oznaczony przez system graficzny jako czerwony nr 172. Porównałby obie czerwienie, *odejmując 163 od 172 i uzyskując 9*, co zinterpretowałby jako dowód, że, powiedzmy, czerwień płaszcza świętego Mikołaja wydaje się odrobinę głębsza i bogatsza (*jemu*) niż czerwień amerykańskiej flagi.

Ta historia została celowo przesadnie uproszczona, aby udramatyzować następujące stwierdzenie: Jest oczywiste, że CAD Dla Niewidomych 1.0 nie używa wymysłu, aby odzwierciedlić swoją pamięć (lub swoją obecną percepcję), *ale my też tego nie robimy*. System ten prawdopodobnie nie wie, jak porównuje barwy czegoś oglądanego z barwami czegoś z pamięci, i *my też tego nie wiemy*. Jest on – przyznam to – raczej prostą, zubożałą przestrzenią barwną z niewieloma skojarzeniami czy wbudowanymi uprzedzeniami barwnej, prywatnej przestrzeni istoty ludzkiej, ale pomimo tej ogromnej różnicy w dyspozycyjnej złożoności nie ma tu różnicy istotnej. Mógłbym nawet powiedzieć tak: nie istnieje *jakościowa* różnica pomiędzy wykonaniem takiego zadania przez system CAD Dla Niewidomych i przez nas. Stany rozróżniania tego systemu mają treść w taki sam sposób i z tych samych powodów jak stany rozróżniania w mózgu, którymi zastąpiłem idee Locke'a. System CAD Dla Niewidomych 1.0 *z pewnością* nie ma żadnych *qualiów* (w tym momencie oczekuję poruszenia ze strony

miłośników *qualiów*), więc rzeczywiście wynika z mojego porównania, że twierdzę, iż my również nie mamy *qualiów*. Ta wyobrażana między każdą maszyną oraz każdym obserwatorem-człowiekiem (przypomnijmy sobie maszynę do smakowania wina, którą wyobrażaliśmy sobie w rozdziale 2) różnica jest z rodzaju tych, jakim niezachwianie zaprzeczam. Nie ma takiej różnicy. Jedynie wydaje się, że jest.

2. Dlaczego istnieją barwy?

Gdy Otto w rozdziale 11 osądził, że wydawało mu się, iż jest tam różowawy, żarzący się pierścień, jaka była treść jego osądu? Jeśli, jak dotychczas twierdziłem, jego osąd nie dotyczył *qualiów*, własności „fenomenalnego", wydającego się pierścienia (zrobionego z wymysłu), o co w nim chodziło? Jaką własność był skuszony przypisać (błędnie) czemuś w świecie zewnętrznym?

Wielu zauważyło, że jest osobliwie trudno powiedzieć, jakimi własnościami obiektów w świecie miałyby być barwy. Prosty i atrakcyjny pomysł – nadal napotykany w wielu elementarnych rozważaniach – jest taki, że każda barwa może być powiązana z niepowtarzalną długością fali świetlnej, a stąd właściwość bycia czerwonym jest po prostu właściwością odbijania całego światła o czerwonej długości fali i absorbowania wszystkich innych jego długości. Jednak od jakiegoś czasu wiadomo już, że nie jest to prawdą. Powierzchnie o różnych fundamentalnych właściwościach odbijania światła mogą być postrzegane jako mające tę samą barwę, a ta sama powierzchnia w innych warunkach oświetlenia może być postrzegana w innych barwach. Długości fal świetlnych wpadających do oka są jedynie pośrednio związane z barwami, w jakich widzimy obiekty. (W Gouras 1984, Hilbert 1987 oraz Hardin 1988 znajdziesz przegląd szczegółów z naciskiem na różne elementy). Dla tych, którzy mieli nadzieję, że znajdzie się prosty, elegancki sposób na wykorzystanie obiecującej uwagi Locke'a o dyspozycyjnych mocach powierzchni, sytuacja nie może być już chyba bardziej niewesoła. Niektórzy (np. Hilbert 1987) zdecydowali się zakotwiczyć kolor obiektywnie, deklarując, że jest on stosunkowo prostą własnością obiektów zewnętrznych, jak właściwość „powierzchniowego odbicia widmowego"; dokonując tego wyboru, muszą następnie dojść do wniosku, że normalne widzenie barwne często podsuwa nam iluzje, skoro postrzegane przez nas stałości tak kiepsko pasują do stałości współczynnika odbicia widmowego mierzonego instrumentami naukowymi. Inni stwierdzili, że właściwości barwne najlepiej rozważać subiektywnie, jako właściwości definiowane ściśle w kategoriach stanów mózgowych obserwatorów, ignorując zagmatwane różnice w świecie, które wywołują te stany: „Barwne obiekty to iluzje, jednak nie bezzasadne. Zwykle znajdujemy się w stanach chromatycznie percepcyjnych, a są to stany neuronalne" (Hardin 1988, s. 111; w Thompson, Palacios i Varela 1991 znajdziesz krytyczne uwagi pod adresem tych poglądów oraz dalsze argumenty na rzecz poglądu przyjmowanego tutaj).

Nie podlega dyskusji, że nie ma prostej, opisywanej bez alternatywy logicznej, własności powierzchni, która odpowiadałaby za to, że wszystkie i jedynie powierzchnie ją mające są czerwone (w sensie cech wtórnych Locke'a). Jest to z początku zastanawiający, a nawet przygnębiający fakt, gdyż wydaje się wskazywać, że nasze percepcyjne rozumienie świata jest znacznie słabsze niż nam się wydawało – że żyjemy w swego rodzaju świecie snów czy też jesteśmy ofiarami masowego złudzenia. Nasze widzenie barwne nie daje nam dostępu do zwykłych właściwości obiektów, nawet jeśli wydaje nam się, że jest inaczej. Dlaczego tak jest?

Czy to po prostu pech? Konstrukcja z niższej półki? Wcale nie. Istnieje inna, o wiele bardziej pouczająca perspektywa, z której możemy spojrzeć na kolor, a została mi ona pokazana po raz pierwszy przez filozofkę neuronauki, Kathleen Akins (1989, 1990)[117]. Czasem nowe

właściwości pojawiają się z jakiegoś powodu. Szczególnie przydatny przykład to słynny przypadek Juliusa i Ethel Rosenbergów, którzy zostali skazani i straceni w 1953 roku za przekazywanie informacji o projekcie amerykańskiej bomby atomowej Związkowi Radzieckiemu. Podczas ich procesu okazało się, że w którymś momencie wpadli na sprytny system haseł: kartonowe pudełko po galaretce było przedzierane na pół, a dwa kawałki przekazywane były dwóm osobom, które musiały być bardzo ostrożne we wzajemnej identyfikacji. Każdy przedarty kawałek stawał się praktycznie niezawodnym i unikatowym „wykrywaczem" partnera: przy późniejszym spotkaniu każdy z nich mógł wyciągnąć swój kawałek i jeśli idealnie się one ze sobą składały, wszystko było w porządku. Dlaczego taki system działa? Ponieważ przedarcie kartonika na pół tworzy krawędź o takiej informacyjnej złożoności, że byłaby właściwie niemożliwa jej celowa reprodukcja. (Zauważmy, że przecięcie kartonika po galaretce ostrą żyletką zupełnie mijałoby się z celem). Konkretna, poszarpana krawędź jednego kawałka staje się *praktycznie* unikatowym urządzeniem wykrywającym wzorzec u partnera; jest to aparat czy przetwornik do wykrywania właściwości kształtu *M*, gdzie *M* jest unikatowo obecny u partnera.

Innymi słowy, właściwość kształtu *M* oraz detektor właściwości *M*, który ją wykrywa, są dla siebie stworzone. Nie byłoby racji istnienia żadnego z nich, stworzenia żadnego z nich w razie nieobecności drugiego. A to samo dotyczy kolorów i widzenia barwnego: zostały dla siebie stworzone. Kodowanie barwami jest dosyć niedawnym pomysłem ergonomii, ale jego walory są już szeroko rozpoznane. Szpitale rozmieszczają barwne linie na korytarzach, ułatwiając pacjentom dotarcie do celu: „Aby dostać się na oddział fizjoterapii, po prostu idź wzdłuż żółtej linii; aby dostać się do banku krwi, idź wzdłuż linii czerwonej!". Producenci telewizorów, komputerów i innych urządzeń elektronicznych kodują barwami duże wiązki kabli wewnątrz w taki sposób, że można łatwo za nimi podążać z miejsca na miejsce. Są to najnowsze zastosowania, ale oczywiście idea jest o wiele starsza; starsza niż szkarłatna litera, którą można było oznaczyć cudzołożnice, starsza niż barwne mundury służące odróżnianiu przyjaciół od wrogów na polu bitwy, a tak naprawdę starsza niż gatunek ludzki.

Zwykle kodowanie barwami pojmujemy jako sprytne wprowadzenie „konwencjonalnych" systemów barw zaprojektowanych tak, aby wykorzystać „naturalne" widzenie barwne, jednak nie bierzemy wówczas pod uwagę faktu, że „naturalne" widzenie barwne wyewoluowało *od samego początku* z nazw, których racją bytu było kodowanie barwami (Humphrey 1976). Niektóre rzeczy w naturze „muszą zostać zobaczone", a inne muszą je zobaczyć, a zatem wyewoluował system, który starał się minimalizować zadania tych drugich przez wzmacnianie wyrazistości tych pierwszych. Przyjrzyjmy się owadom. Ich widzenie barwne wyewoluowało równolegle z barwami roślin, które zapylały – świetny trik konstrukcyjny, który dał zysk obu stronom. Bez kodowania barwą kwiatów widzenie barwne u owadów nie wyewoluowałoby, i na odwrót. Zatem zasadą kodowania barwą jest podstawowe widzenie barwne u owadów, a nie tylko niedawne wynalazki jednego bystrego gatunku ssaków. Podobne historie można opowiedzieć o ewolucji widzenia barwnego u innych gatunków. Podczas gdy pewien rodzaj tego widzenia mógł początkowo wyewoluować z wizualnego rozróżniania zjawisk nieorganicznych, cały czas nie jest jasne, czy stało się to u jakiegokolwiek gatunku na naszej planecie. (Evan Thompson zwrócił moją uwagę na to, że pszczoły mogą korzystać ze szczególnego rodzaju widzenia barwnego w nawigacji, aby rozróżniać spolaryzowane światło słoneczne w pochmurne dni. Czy jest to jednak wtórne wykorzystanie widzenia barwnego, które początkowo pojawiło się równolegle z barwami kwiatów?)

Różne systemy widzenia barwnego wyewoluowały niezależnie, czasem z radykalnie różnymi przestrzeniami barw. (Krótki przegląd oraz odwołania do tej kwestii znajdziesz

w Thompson, Palacios i Varela 1991). Nie wszystkie stworzenia z oczami mają jakieś widzenie barwne. Ptaki, ryby, gady i owady wyraźnie są w nie wyposażone, a jest ono podobne do naszego systemu „trójchromatycznego" (czerwony, zielony, niebieski); psy i koty go nie mają. Wśród ssaków tylko naczelne są wyposażone w widzenie barwne, ale są pomiędzy nimi szokujące różnice. Które gatunki mają widzenie barwne i dlaczego? Okazuje się to fascynującą i skomplikowaną historią, która nadal w dużej mierze opiera się na spekulacjach.

Dlaczego jabłka stają się czerwone, gdy dojrzewają? Naturalne jest przyjąć, że cała odpowiedź może zostać udzielona w kategoriach chemicznych zmian zachodzących w momencie, gdy cukier i inne substancje osiągają pewne stężenie w dojrzewającym owocu, prowadząc do różnych reakcji itd. Jednak byłoby to zignorowaniem faktu, że w ogóle nie byłoby jabłek, gdyby nie istniały jedzące je i roznoszące nasiona osobniki, więc fakt, iż jabłka są wyraźnie widoczne dla przynajmniej niektórych ich konsumentów, jest warunkiem ich istnienia, a nie zwykłym „przypadkiem" (z punktu widzenia jabłka!). Fakt, że jabłka mają takie, a nie inne widmowe właściwości odbijania powierzchniowego, jest tak samo funkcją fotopigmentów, które można było wykorzystać w czopkach w oku konsumentów owoców, jak efektem interakcji między cukrem i innymi związkami w chemii owocu. Owoce pozbawione barw nie są konkurencją na półkach supermarketu natury, jednak nieprawdziwa reklama skończy się karą; dojrzałe owoce (pełne składników odżywczych), *które reklamują ten fakt*, będą się sprzedawać lepiej, lecz reklama musi być dopasowana do wizualnych możliwości i skłonności docelowych konsumentów.

Na początku barwy powstały, aby widzieli je ci, którzy mieli je oglądać. Jednak stało się to stopniowo, przypadkiem, z nieoczekiwanym wykorzystaniem materiałów, które akurat były pod ręką, okazjonalnie wybuchając obfitością przeróbek nowej sztuczki i zawsze tolerując sporą ilość bezcelowych wariacji *oraz* bezcelowej (zaledwie przypadkowej) stałości. Te przypadkowe stałości często dotyczyły „bardziej fundamentalnych" cech świata fizycznego. Gdy już pojawiły się istoty mogące odróżnić jagody czerwone od zielonych, mogły one również odróżnić czerwone rubiny od zielonych szmaragdów, ale była to jedynie przypadkowa premia. Fakt, że istnieje różnica *w barwie* między rubinami i szmaragdami może zatem być uważany za *pochodne* zjawisko barwy. Dlaczego niebo jest niebieskie? Ponieważ jabłka są czerwone, a śliwki fioletowe, nie na odwrót.

Błędem jest myślenie, że pierwsze były barwy – barwne skały, barwna woda, barwne niebo, czerwonawa rdza i jasnoniebieski kobalt – a potem przyszła Matka Natura i wykorzystała *te* właściwości, kodując wszystko barwami. Pierwsze były raczej różne właściwości odbijania od powierzchni, reaktywne właściwości fotopigmentów itd., a Matka Natura rozwinęła z tych surowych materiałów wydajne, wzajemnie dostrojone kodujące „barwami"/„barwne" systemy widzenia, zaś pośród cech powstałych w tym procesie projektowania są takie, które my, istoty ludzkie, nazywamy „barwami". Jeśli niebieski kobaltu i niebieski skrzydła motyla są takie same (dla normalnego widzenia ludzkiego), jest to tylko zbieg okoliczności, nieistotny skutek uboczny procesów, które stworzyły widzenie barwne, a przez to (jak mógłby przyznać sam Locke) ochrzciły pewien niezwykle dziwacznie zbudowany zestaw złożoności cech pierwotnych wspólną cechą wtórną, wywołującą wspólny efekt w grupie normalnych obserwatorów.

„Ale przecież – będziesz oponować – gdy nie było jeszcze żadnych zwierząt z widzeniem barwnym, istniały wspaniałe zachody słońca i błyszczące, zielone szmaragdy!" Cóż, można tak powiedzieć, ale z drugiej strony te same zachody słońca były również jaskrawe, wielokolorowe i obrzydliwe, mając barwy, których nie widzimy, więc nie mamy na nie nazwy. To znaczy, musisz jednak przyznać, jeśli istnieją *lub mogłyby istnieć* istoty na jakiejś planecie, których aparat sensoryczny byłby w taki sposób przez te zachody rozbudzony. A o ile nam wiadomo,

istnieją gdzieś gatunki, które w sposób naturalny widzą, że są dwie (lub siedemnaście) barwy pośród garści szmaragdów, które *my* uznaliśmy za bezdyskusyjnie zielone.

Wielu ludzi nie rozpoznaje barwy czerwonej. Przypuśćmy, że nie rozpoznajemy jej wszyscy; wówczas byłoby powszechnie wiadomo, że rubiny i szmaragdy są „zieloczerwone" – w końcu dla zwykłych obserwatorów wyglądają jak inne zieloczerwone przedmioty: samochody strażackie, odpowiednio podlewane trawniki, dojrzałe i niedojrzałe jabłka (Dennett 1969). Gdybyśmy wtedy stwierdzili, że rubiny i szmaragdy mają inny kolor, nie sposób byłoby stwierdzić, że jeden z tych systemów widzenia barw jest „prawdziwszy" od drugiego.

Filozof Jonathan Bennett (1965) zwraca uwagę na przypadek dowodzący tego samego, choć bardziej przekonująco, w przypadku innej modalności sensorycznej. Mówi, że substancja fenylotiokarbamid jest gorzka dla jednej czwartej populacji ludzkiej, dla reszty będąc zupełnie bez smaku. Jej smak jest zdeterminowany genetycznie. Czy fenylotiokarbamid jest gorzki, czy bez smaku? Poprzez „eugenikę" (kontrolowane rozmnażanie) czy też inżynierię genetyczną może udać nam się eliminacja genotypu powodującego, że smakuje on gorzko. Gdyby się nam to udało, fenylotiokarbamid byłby wówczas *paradygmatycznie* bez smaku, jak woda destylowana: bez smaku dla wszystkich normalnych istot ludzkich. Gdybyśmy przeprowadzili odwrotny eksperyment genetyczny, moglibyśmy sprawić, że z czasem substancja ta stałaby się paradygmatycznie gorzka. Zanim pojawiły się istoty ludzkie, czy fenylotiokarbamid był *zarówno* gorzki, jak i bez smaku? Chemicznie był taki sam jak teraz.

Fakty dotyczące właściwości wtórnych są nierozerwanie powiązane z klasą odniesienia obserwatorów, lecz istnieją słabe i silne podejścia do tego połączenia. Możemy powiedzieć, że właściwości wtórne są *śliczne*, ale nie *podejrzane*. Śliczny mógłby być ktoś, kto jeszcze nigdy nie był obserwowany przez obserwatora w rodzaju tego, który uznałby go za ślicznego, jednak nie mógłby – logicznie rzecz biorąc – być podejrzany, dopóki ktoś rzeczywiście by go o coś podejrzewał. O konkretnych przykładach jakości ślicznych (jak cecha śliczności) można powiedzieć, że istnieją jako dyspozycje Locke'a przed momentem (jeśli w ogóle), w którym objawiają moc nad obserwatorem, wywołując w nim określony efekt. Zatem jakaś niewidziana kobieta (która dorastała sama na bezludnej wyspie) mogłaby być prawdziwie śliczna, mając dyspozycyjną moc wpływania na normalnych obserwatorów pewnego rodzaju w pewien sposób, mimo że nigdy nie miała możliwości, aby tego dokonać. Jednak jakości śliczności nie mogą być zdefiniowane niezależnie od skłonności, wrażliwości czy dyspozycji jakiejś klasy obserwatorów, więc nie ma tak naprawdę sensu mówienie o istnieniu cechy śliczności w całkowitym oderwaniu od istnienia odpowiednich obserwatorów. Choć jest to za mocno powiedziane. Właściwość śliczności *nie byłaby* definiowana – nie byłoby sensu *jej* definiowania, w przeciwieństwie do wszystkich innych, logicznie możliwych, choćby dziwacznie wyodrębnionych cech – niezależnie od takiej klasy obserwatorów. A zatem, choć mogłoby być logicznie możliwe (można by powiedzieć: „z perspektywy czasu") zebranie przykładów właściwości barwnych przez coś w rodzaju zwykłego i wyczerpującego wyliczenia, powody wyróżnienia takich cech (na przykład w celu wyjaśnienia pewnych przyczynowych prawidłowości w zbiorze dziwnie złożonych obiektów) zależą od istnienia takiej klasy obserwatorów.

Czy słonie morskie są śliczne? Nie dla nas. Trudno wyobrazić sobie brzydszą istotę. To, co sprawia, że słoń morski jest śliczny dla innego słonia morskiego, to nie to samo, co sprawia, że kobieta jest śliczna dla mężczyzny, a nazwanie śliczną nadal niezaobserwowanej kobiety, która – tak się składa – niezmiernie podoba się słoniom morskim, byłoby nadużyciem zarówno w odniesieniu do niej, jak i tego pojęcia. Jedynie przez odwołanie do ludzkich gustów, przypadkowych i bardzo dziwacznych cech w świecie, właściwość bycia ślicznym (dla istoty ludzkiej) może zostać zidentyfikowana.

Z drugiej strony jakość bycia podejrzanym jest rozumiana w taki sposób, aby zakładać, że każda egzemplifikacja tej cechy wywarła określony skutek na przynajmniej jednego obserwatora. Być może szczególnie warto cię podejrzewać – twoja wina może być wręcz oczywista – jednak nie można być podejrzanym lub podejrzaną, dopóki ktoś rzeczywiście nie zacznie nas podejrzewać. Nie twierdzę, że kolory są podejrzanymi jakościami. Nie musimy zaprzeczać naszemu przeczuciu, że jeszcze niezaobserwowany szmaragd wewnątrz grudki w rudzie *już* jest zielony. Twierdzę jednak, że kolory to jakości śliczne, których istnienie powiązane z klasą odniesienia obserwatorów nie ma sensu w świecie, w którym nie ma obserwatorów. Jest to łatwiejsze do przyjęcia w przypadku pewnych cech wtórnych niż innych. To, że opary siarki wyrzucane przez pradawne wulkany były jakoś żółte, wydaje się bardziej obiektywne niż to, że cuchnęły, ale dopóki to, co rozumiemy przez „żółte", jest tym, co *my* rozumiemy przez „żółte", te twierdzenia są analogiczne. Przyjmijmy, że jakieś pradawne trzęsienie ziemi utworzyło przepaść odsłaniającą pasy setek chemicznie odmiennych warstw atmosfery. Czy te pasy były *widzialne*? Musimy zapytać dla kogo. Być może niektóre z nich byłyby dla nas widoczne, a inne nie. Być może niektóre z niewidzialnych pasów byłyby widoczne dla tetrachromatycznych gołębi lub dla stworzeń widzących podczerwony bądź ultrafioletowy zakres promieniowania elektromagnetycznego. Z tego samego powodu, dla którego nie można sensownie zapytać, czy różnica między szmaragdami i rubinami jest różnicą widoczną, bez określenia, o jaki układ widzenia nam chodzi.

Ewolucja łagodzi cios „subiektywizmu" czy „relatywizmu" sugerowany przez fakt, iż cechy wtórne są jakościami ślicznymi. Pokazuje, że nieobecność „prostych" czy „fundamentalnych" cech wspólnych przedmiotów tej samej barwy nie świadczy o totalnej iluzji, lecz raczej o szeroko rozpowszechnionej tolerancji „fałszywie pozytywnych" detekcji ekologicznych właściwości, które rzeczywiście mają znaczenie[118]. Podstawowe kategorie naszej przestrzeni barw (a także oczywiście przestrzeni zapachów i dźwięków oraz wszystkich innych) są ukształtowane przez presje selekcyjne, więc – ogólnie rzecz biorąc – można zadać pytanie, czemu służy konkretne rozróżnienie czy preferencja. Istnieją racje, dla których odrzucamy zapachy pewnych rzeczy, a wyciągamy ręce do innych, dla których wolimy jedne barwy od drugich, dla których pewne dźwięki bardziej nas denerwują lub bardziej uspokajają. Nie muszą one być zawsze *naszymi* racjami, a mogą raczej być racjami dalekich przodków, którzy pozostawili swoje ślady we wrodzonych skłonnościach z natury kształtujących naszą przestrzeń *qualiów*. Ale jako dobrzy następcy Darwina powinniśmy również uznać możliwość – a nawet konieczność – innych, niefunkcjonalnych skłonności, przypadkowo rozłożonych w zmiennej genetycznie populacji. Aby presja selekcyjna różnorako wspierała tych preferujących *F*, gdy *F* staje się ekologicznie ważne, musiała zaistnieć bezsensowna (jeszcze niefunkcjonalna) zmiana „stosunku do *F*", na którą mogła działać selekcja. Gdyby na przykład jedzenie flaków miało w przyszłości doprowadzić do reprodukcyjnej zagłady, jedynie ci z nas będący „naturalnie" (i *dotychczas* bezsensownie) skłonni do niejedzenia flaków mieliby przewagę (być może z początku znikomą, ale wkrótce potężną, jeśli wspierałyby ją warunki). Nie oznacza to więc, że jeśli znajdziesz coś nieopisanie i niewyrażalnie obrzydliwego (na przykład brokuły), jest po temu racja. Nie oznacza to również, że jest z tobą coś nie tak, jeśli nie zgadzasz się co do tego z innymi. Może to być tylko jedno z wrodzonych wybrzuszeń w twojej przestrzeni *qualiów*, które jeszcze nie ma żadnego znaczenia funkcjonalnego. (A dla twojego dobra najlepiej byłoby, żeby flaki zyskały na znaczeniu właśnie dlatego, że brokuły okazałyby się dla nas szkodliwe).

Te rozważania ewolucyjne walnie przyczyniają się do wyjaśnienia, dlaczego cechy wtórne okazują się tak „niewyrażalne", tak opierające się definicji. Jak właściwość kształtu *M* fragmentu kartonika po galaretce Rosenbergów, cechy wtórne są wyjątkowo odporne na proste

definicje. Esencją triku Rosenbergów jest to, że nie możemy wymienić naszego fikcyjnego predykatu *M* na dłuższy, bardziej złożony, ale dokładny i wyczerpujący opis tej cechy, bo gdybyśmy byli w stanie to zrobić, my (lub ktoś inny) mógłby wykorzystać go jako przepis produkcji kolejnej egzemplifikacji *M* lub kolejnego wykrywacza *M*. Nasze detektory cech wtórnych nie były skonstruowane po to, by wykrywać jedynie właściwości trudne do zdefiniowania, ale rezultat jest właśnie taki. Jak zauważa Akins (1989), *celem naszych układów zmysłowych nie jest* wykrywanie „podstawowych" czy „naturalnych" właściwości środowiska, lecz realizowanie naszego „narcystycznego" celu przeżycia; natura nie buduje silników epistemicznych.

Jedynym *łatwo dostępnym sposobem* stwierdzenia, czym jest właściwość kształtu *M*, jest wskazanie detektora *M* i powiedzenie, że *M* to właściwość kształtu przez niego wykrywana. W tak samo trudnym położeniu jest oczywiście każdy, kto stara się powiedzieć, jaką cechę ktoś wykrywa (lub wykrywa błędnie), gdy coś „wygląda tak, jak dla niego wygląda". Zatem teraz możemy odpowiedzieć na pytanie, od którego zaczął się ten podrozdział: Jaką cechę osądza Otto jako przysługującą czemuś, gdy sądzi, że to coś jest różowe? Cechę, którą nazywa różową. A co to za cecha? Trudno powiedzieć, ale nie powinno nas to zawstydzać, gdyż potrafimy stwierdzić, dlaczego trudno to powiedzieć. Praktycznie najlepsze, co możemy zrobić zapytani o to, jakie właściwości powierzchni wykrywamy w widzeniu barwnym, to powiedzieć (niepouczająco), że wykrywamy cechy, które wykrywamy. Jeśli ktoś żąda bardziej pouczającej historii o tych cechach, może zajrzeć do ogromnej i raczej mało zrozumiałej literatury z dziedziny biologii, neuronauki i psychofizyki. A Otto nie może powiedzieć nic więcej o cesze, którą nazywa różową, niż „To jest *to*!" (wskazując „wewnątrz" na prywatną, fenomenalną cechę swojego doświadczenia). Wszystko, co można przez to osiągnąć (w najlepszym razie), to wskazać na jego własne, specyficzne stany rozróżniania barw, co podobne jest do chwycenia fragmentu kartonika po galaretce i powiedzenia, że wykrywa *tę* cechę kształtu. Otto być może wskazuje na swoje urządzenie wykrywające, ale nie na żadne *quale*, które jest przez nie wydzielane lub przez nie noszone, lub przez nie odbijane, gdy ono działa. Coś takiego jak *quale* nie istnieje.

No ale przecież – upiera się Otto – nie powiedziałeś jeszcze, dlaczego różowy wygląda *tak*!

Jak?

Tak. Jak szczególnie niewyrażalna, wspaniała, wewnętrzna różowość, która mnie właśnie teraz cieszy. *To* nie jest jakaś nieopisanie zawiła właściwość odbijania światła od powierzchni przedmiotów zewnętrznych.

Widzę, Otto, że użyłeś terminu „*cieszy*". Nie jesteś sam. Często, gdy autor chce podkreślić, że temat przeszedł z (jedynie) neuroanatomicznego do przeżyć świadomych, z (jedynie) psychofizyki do świadomości, z (jedynie) informacji na *qualia*, na scenę wkracza słówko „cieszy".

3. Radość z naszych doświadczeń

Ale, Dan, *qualia* sprawiają,
że życie warte jest zachodu.
Wilfrid Sellars (przy butelce znakomitego chambertin, Cincinnati, 1971)

Jeśli tym, czego chcę, pijąc dobre wino, jest informacja o jego właściwościach chemicznych, to dlaczego po prostu nie przeczytam etykiety?

Sydney Shoemaker,

seminarium na Uniwersytecie Tufts, 1988

Niektóre barwy powstały po to, by być lubiane, tak jak niektóre zapachy i smaki. A inne barwy, zapachy i smaki powstały po to, by nie być lubiane. Dobierając słowa bardziej delikatnie, nie jest przypadkiem, że my (i inne stworzenia, które je wykrywają) lubimy i nie lubimy barw, zapachów, smaków i innych cech wtórnych. Tak jak odziedziczyliśmy ewolucyjnie detektory pionowej symetrii w układach wzrokowych, mające nas ostrzegać (jak naszych przodków) o ekologicznie istotnym fakcie, że patrzy na nas inne stworzenie, tak też otrzymaliśmy spadek w postaci wykształconych detektorów cech niebędących bezinteresownymi reporterami, a raczej detektorami, które ostrzegają i wabią, syrenami zarówno w sensie samochodu strażackiego, jak i w sensie homeryckim.

Jak widzieliśmy w rozdziale 7 o ewolucji, ci rodzimi alarmiści zostali następnie wchłonięci przez mnóstwo bardziej skomplikowanych organizacji zbudowanych z milionów połączeń i ukształtowanych, w przypadku ludzi, przez tysiące memów. W ten sposób brutalne wezwanie „bierz mnie" dotyczące seksu i jedzenia oraz niemiłosierna niechęć do bólu i strachu polegająca na ciągłej ucieczce mieszają się razem w przeróżne rodzaje pikantnych kombinacji. Gdy organizm odkrywa, że opłaca mu się zajmować jakąś cechą w świecie *pomimo* wbudowanej do tego awersji, musi skonstruować równoważącą koalicję, aby awersja nie wygrała. Wynikające z tego częściowo stabilne napięcie może następnie stać się nabytym gustem, poszukiwanym w pewnych okolicznościach. Gdy organizm odkrywa, że musi tłumić efekty pewnych natarczywych kusicieli, jeśli chce przyjąć zamierzony kurs, może wyrobić w sobie gust na wszelkie sekwencje czynności, jakie tylko są możliwe, aby mieć skłonność do tworzenia pożądanych: ciszy i spokoju. W ten sposób mogliśmy zacząć uwielbiać ostre jedzenie, palące nasze usta (Rozin 1982), wybornie „nieharmonijną" muzykę oraz zarówno opanowany realizm Andrew Wyetha, jak i niepokojący, gorący ekspresjonizm Willema de Kooninga. Marshall McLuhan (1967) głosił, że środek przekazu sam jest przekazem, co jest półprawdą być może prawdziwszą w systemie nerwowym niż w jakiejkolwiek innej formie komunikacji. To, czego pragniemy, popijając dobre wino, to rzeczywiście nie informacja o jego składzie chemicznym; chcemy być *poinformowani* o jego zawartości chemicznej w nasz ulubiony sposób. A nasza preferencja jest *ostatecznie* oparta na uprzedzeniach nadal wbudowanych w układy nerwowe, choć ich znaczenie ekologiczne mogło zaniknąć wieki temu.

Ten fakt jest przed nami w dużej mierze skrywany przez naszą własną technikę. Jak ujął to psycholog Nicholas Humphrey:

Gdy patrzę na pokój, w którym pracuję, barwa stworzona przez człowieka krzyczy na mnie ze wszystkich stron: książki, poduszki, czerwienie, żółcie, zielenie. Jest tu tyle barw, ile w lesie tropikalnym. Jednak w lesie każda barwa miałaby swoje znaczenie, a tu, w moim gabinecie, żadna go nie ma. Władzę przejęła kolorystyczna anarchia. [Humphrey 1983, s. 149]

Przyjrzyjmy się na przykład dziwnemu faktowi, że małpy nie lubią czerwonego światła. Mając możliwość wyboru, rezusy wykazują silną preferencję do niebiesko-zielonej części spektrum, a denerwują się, musząc spędzać czas w środowisku czerwonym (Humphrey 1972, 1973, 1983; Humphrey i Keeble 1978). Dlaczego tak jest? Humphrey zauważa, że czerwień zawsze jest używana jako ostrzeżenie, ostateczna barwa w kodowaniu barwami, ale właśnie z tego powodu jest dwuznaczna: czerwony owoc może być dobry do zjedzenia, ale czerwony wąż czy owad prawdopodobnie reklamuje się jako jadowity. Więc „czerwony" przesyła niejasną wiadomość. Po co jednak w ogóle wysyła wiadomość „alarm"? Być może dlatego, że jest najsilniejszym dostępnym kontrastem wobec otaczającego tła, roślinnego zielonego lub morskiego niebieskiego, lub – jak w przypadku małp – ponieważ czerwone światło (od

czerwonego przez czerwono-pomarańczowe aż po pomarańczowe) jest światłem zmierzchu i świtu, chwili w ciągu dnia, gdy polują niemalże wszystkie drapieżniki atakujące małpy.

Afektywne czy emocjonalne cechy czerwieni nie ograniczają się do rezusów. Wszystkie naczelne mają takie reakcje, w tym również ludzie. Jeśli pracownicy twojego zakładu zbyt dużo spędzają na pogaduszkach w toaletach, pomalowanie ścian na czerwono rozwiąże problem – ale stworzy kolejne (zob. Humphrey 1992). Takie „wewnętrzne" reakcje nie ograniczają się oczywiście do barw. Większość naczelnych chowanych w niewoli, które nigdy nie widziały węża, gdy tylko go zobaczy, bardzo wyraźnie da do zrozumienia, że ich nie znosi, i jest bardzo prawdopodobne, że tradycyjna ludzka niechęć do węży ma źródło biologiczne, które wyjaśnia źródła biblijne, a nie na odwrót[119]. To znaczy, że nasze genetyczne dziedzictwo działa na korzyść memów odpowiadających za nienawiść do węży.

Oto dwa różne wyjaśnienia dla niepokoju (nawet gdy go „opanujemy"), jaki wielu z nas czuje, gdy widzi węża:

(1) Węże wywołują w nas szczególne, wewnętrzne *quale* ohydy na widok węża, a nasz niepokój jest na nie reakcją.

(2) Nie chcemy patrzeć na węże z powodu wewnętrznych skłonności wbudowanych w nasz układ nerwowy. Powodują one uwolnienie adrenaliny, prowadzą do reakcji „uciekaj bądź walcz" i poprzez aktywowanie przeróżnych połączeń przywołują mnóstwo scenariuszy związanych z zagrożeniem, przemocą, zniszczeniem. Pierwotna awersja naczelnych jest w nas przemieniona, skorygowana, odwrócona na setki sposobów przez memy ją wykorzystujące, wchłaniające, kształtujące. (Jest wiele różnych poziomów, na których moglibyśmy sformułować wyjaśnienie w takim „funkcjonalistycznym" typie. Na przykład moglibyśmy sobie pozwolić na bardziej pobieżne mówienie o mocy percepcji węży do wywoływania stresu, strachu, przewidywanego bólu itp., jednak mogłoby to zostać odebrane jako „oszukiwanie", więc tego unikam).

Problem z pierwszym wyjaśnieniem jest taki, że wydaje się jedynie wyjaśnieniem. Pomysł, że „wewnętrzna" własność (bieżącego różowego, ohydy węża, bólu, aromatu kawy) mogłaby *wyjaśnić* reakcje podmiotu na jakąś okoliczność, jest fatalny – oczywisty przypadek *virtus dormitiva* (zob. rozdz. 3). Uznanie teorii za skrywającą bezmyślną *virtus dormitiva* nie jest jednak takie proste. Czasem ma sens przyjęcie tymczasowej *virtus dormitiva*, oczekującej na dalsze rozpatrzenie. Poczęcie, moglibyśmy powiedzieć, jest z definicji przyczyną ciąży. Gdyby nie można było inaczej zdefiniować poczęcia, powiedzenie jakiejś osobie, że zaszła w ciążę, gdyż doszło do poczęcia, byłoby pustym gestem, a nie wyjaśnieniem. Gdy jednak wypracujemy wymaganą, mechaniczną teorię poczęcia, widzimy, *jak* jest ono przyczyną ciąży, i informacja zostaje przywrócona. W ten sam sposób moglibyśmy identyfikować *qualia*, z definicji, jako bliższe przyczyny naszej radości i cierpienia (mówiąc z grubsza), a następnie wypełniać obowiązek informowania i dążyć do wyjaśnienia typu drugiego. Ale co ciekawe, qualofile (jak nazywam tych, którzy nadal wierzą w *qualia*) tego nie zaakceptują; twierdzą, jak Otto, że *qualia* „zredukowane" do zwykłych zespołów mechanicznie osiąganych dyspozycji do reagowania nie są *qualiami*, o których mówią. *Ich qualia* są czymś innym.

Pomyśl – mówi Otto – jaki wydaje mi się *teraz* różowy pierścień, w tym momencie, w izolacji od wszelkich moich dyspozycji, przeszłych połączeń i przyszłych działań. *Ten* czysty, wyizolowany *sposób, w jaki mi się on wydaje*, jeśli chodzi o barwę w tym momencie – to moje różowe *quale*.

Otto właśnie popełnił błąd. Jest to tak naprawdę duży błąd, źródło wszelkich paradoksów

związanych z *qualiami*, jak wkrótce zobaczymy. Ale zanim odkryjemy, dlaczego niemądrze jest iść tą ścieżką, chcę zademonstrować pewne zalety drogi, którą odrzuca Otto: drogi „redukcjonistycznej", polegającej na *utożsamieniu* „tego, jak to jest ze mną" z sumą wszystkich indywidualnych dyspozycji do reagowania, wbudowanych w mój układ nerwowy w wyniku mojej konfrontacji z pewnym wzorcem pobudzeń.

Pomyślmy, czym musiało być dla luteranina z Lipska w roku, powiedzmy, 1725 słuchanie jednej z kantat Jana Sebastiana Bacha w premierowym wykonaniu. (Ćwiczenie w wyobrażaniu sobie, *jak to jest*, stanowi rozgrzewkę przed rozdziałem 14, gdzie zajmiemy się świadomością u innych zwierząt). Nie istnieją prawdopodobnie żadne biologiczne różnice pomiędzy nami, żyjącymi dziś, a niemieckimi luteranami z XVIII wieku; jesteśmy tym samym gatunkiem i nie dzieli nas wiele czasu. Jednak z powodu potężnego wpływu kultury – memosfery – nasz świat psychiczny jest nieco inny od ich świata, pod względem istotnym dla przeżycia słuchania kantaty Bacha po raz pierwszy. Nasza wyobraźnia muzyczna została wzbogacona oraz skomplikowana na wiele sposobów (przez Mozarta, Charliego Parkera, Beatlesów), ale również straciła pewne potężne skojarzenia, na które mógł liczyć Bach. Jego kantaty były zbudowane na chorałach, melodiach z tradycyjnych pieśni, dobrze znanych uczęszczającym do kościoła i dlatego wywołujących fale emocjonalnych i tematycznych skojarzeń, gdy tylko ich ślady i echa pojawiały się w muzyce. Większość z nas zna te chorały tylko z dzieł Bacha, więc gdy je słyszymy, robimy to innymi uszami. Jeśli chcemy sobie wyobrazić, jak to było słuchać Bacha w osiemnastowiecznym Lipsku, nie wystarczy posłuchać tych samych dźwięków na tych samych instrumentach w tej samej kolejności; musimy także jakoś przygotować się, aby odpowiedzieć na te dźwięki z tym samym smutkiem, dreszczem i falą nostalgii.

Takie przygotowanie nie jest zupełnie niemożliwe. Muzykolog, który ostrożnie unikał wszelkiego kontaktu z muzyką stworzoną po roku 1725 i dogłębnie zapoznał się z tradycyjną muzyką tamtego okresu, znacznie by się do tego celu przybliżył. Co ważniejsze, jak pokazują te obserwacje, nie jest niemożliwe wiedzieć, jak musielibyśmy się przygotować, bez względu na to, czy chciałoby się nam podejmować ten trud. Moglibyśmy więc wiedzieć, jak to jest, w sposób, że tak powiem, „abstrakcyjny", a tak naprawdę lipszczanom słuchającym kantat przychodziły do głowy wszystkie skojarzenia, które nadawały smak ich odbiorowi melodii chorałów. Łatwo jest sobie wyobrazić, jakie *to* musiało dla nich być – choć z różnicami zapożyczonymi z naszego własnego doświadczenia. Możemy sobie wyobrazić, jak by to na przykład było usłyszeć stworzoną przez Bacha wersję znanych nam kolęd czy *Home on the Range*. Nie możemy tego zrobić dokładnie, ale tylko dlatego, że nie potrafimy zapomnieć czy porzucić wszystkiego, co wiemy, a o czym nie wiedzieli lipszczanie.

Aby zobaczyć, jak istotny jest dla nas ten bagaż, wyobraźmy sobie, że muzykolodzy dotarli do dotychczas nieznanej kantaty Bacha, z pewnością przez niego stworzonej, ale schowanej do szuflady i prawdopodobnie nigdy nie słyszanej, nawet przez samego kompozytora. Wszyscy chcieliby ją usłyszeć, po raz pierwszy doświadczyć „*qualiów*", jakie poznaliby lipszczanie, gdyby tylko ten utwór usłyszeli, ale okazuje się to niemożliwe, gdyż główna melodia kantaty, w wyniku szpetnego zbiegu okoliczności, to siedem pierwszych dźwięków *Rudolfa Czerwononosego*! Obciążeni tą melodią, *nigdy* nie bylibyśmy w stanie wysłuchać wersji Bacha, jak zakładał to sam kompozytor, albo tak, jak odebraliby go lipszczanie.

Trudno byłoby znaleźć bardziej oczywisty przypadek blokady wyobraźni, ale zauważmy, że nie ma on nic wspólnego z różnicami biologicznymi ani nawet z „wrodzonymi" czy „niewyrażalnymi" właściwościami muzyki Bacha. Powód, dla którego nie moglibyśmy szczegółowo (i odpowiednio) przeżyć w wyobraźni muzycznego przeżycia, które przeżywali lipszczanie, jest po prostu taki, że musielibyśmy zabrać samych siebie w tę wyimaginowaną

podróż, a wiemy zbyt wiele. Jeśli jednak chcemy, możemy zrobić dokładną listę różnic między naszymi dyspozycjami i wiedzą a ich dyspozycjami i wiedzą oraz przez porównanie tych list docenić, *tak dokładnie, jak tylko mamy ochotę*, różnice między tym, czym słuchanie Bacha było dla nich i czym jest dla nas. Moglibyśmy lamentować nad trudnym dostępem do takiego przeżycia, ale przynajmniej zrozumielibyśmy je. Nie pozostałaby żadna tajemnica; tylko przeżycie, które można by dosyć dokładnie opisać, ale z którego nie można by bezpośrednio się cieszyć, chyba że podjęlibyśmy się absurdalnego zadania przebudowania naszych osobistych struktur dyspozycyjnych.

Qualofile odrzucają jednak ten wniosek. Wydaje im się, że chociaż badanie właśnie przez nas wyobrażone mogłoby odpowiedzieć na *prawie* wszystkie pytania dotyczące tego, jak to jest być lipszczaninem, musiałby pozostać jakiś niewyrażalny osad, *coś* dotyczące tego, jak to jest być lipszczaninem, czego żadne dalsze postępy w zaledwie „dyspozycyjnej" i „mechanicznej" wiedzy nie mogłyby zredukować do zera. Właśnie dlatego qualofile muszą przywoływać *qualia* jako właściwości *dodatkowe*, istniejące ściśle niezależnie od okablowania determinującego wycofanie, zmarszczenie czoła, krzyknięcie oraz inne „błahe zachowania" związane z obrzydzeniem, nienawiścią, strachem; ba, wykraczające poza i ponad takie okablowanie. Możemy to jasno zobaczyć, gdy powrócimy do przykładu barw.

Załóżmy, że zasugerowaliśmy Ottonowi, iż to, co sprawiało, że jego „bieżący różowy" był szczególnie wspaniałym przeżyciem, które dawało mu radość, było po prostu sumą wszystkich wrodzonych oraz wyuczonych skojarzeń i dyspozycji do reagowania spowodowanych szczególnymi (błędnymi) informacjami z oczu:

Ottonie, *qualia* to po prostu zespoły dyspozycji. Gdy mówisz „*To* jest moje *quale*", to odwołujesz czy też odnosisz się, *bez względu na to, czy sobie z tego zdajesz sprawę*, do swoich osobistych zespołów dyspozycji. *Wydaje się*, że odnosisz się do prywatnego, niewyrażalnego „czegoś" w swojej wyobraźni, prywatnego odcienia jednolitego różu, ale tylko tak ci się wydawało, a nie było tak w rzeczywistości. Owo „*quale*" to wiarygodna postać w fikcyjnym świecie twojej heterofenomenologii, ale w *prawdziwym* świecie, w twoim mózgu, okazuje się jedynie zespołem dyspozycji.

To nie może być wszystko – odpowiada Otto, kierując się podszeptami fatalnej tradycji qualofilskiej – bo chociaż zespół zwykłych dyspozycji może być jakąś podstawą czy źródłem mojego konkretnego *quale* koloru różowego, cały ten zespół mógłby zostać zmieniony bez zmiany mojego wewnętrznego *quale*, bądź też moje wewnętrzne *quale* mogłoby się zmienić bez zmiany tych rozmaitych dyspozycji. Na przykład moje *qualia* mogłyby zostać *odwrócone* bez odwrócenia wszystkich moich dyspozycji. Mógłbym mieć wszystkie reakcje i skojarzenia, które mam teraz dla barwy *zielonej* razem z *quale*, które mam teraz dla koloru *czerwonego*, i na odwrót.

4. Filozoficzna fantazja: odwrócone *qualia*

Idea możliwości takich „odwróconych *qualiów*" jest jednym z najbardziej złośliwych memów filozofii. Locke zgłębiał ją w swoich *Rozważaniach dotyczących rozumu ludzkiego* (1690/1955), a wielu studentów mówi mi, że jako małe dzieci wpadli na ten sam pomysł i byli nim zafascynowani. Pomysł wydaje się oczywisty i bezpieczny:

Rzeczy mają dla mnie różne wyglądy, brzmienia, zapachy itd. To jest oczywiste. Zastanawiam się jednak, czy to, jakie rzeczy wydają się mi, jest tym samym, czym wydają się innym.

Filozofowie stworzyli wiele różnorodnych wersji tego motywu, choć wersja klasyczna

jest interpersonalna: Skąd wiem, że ty i ja widzimy tę samą *subiektywną* barwę, gdy na coś patrzymy? Skoro nauczyliśmy się nazw barw, gdyż pokazywano nam zewnętrzne, kolorowe obiekty, nasze zachowanie werbalne będzie takie samo, nawet jeśli doświadczamy zupełnie innych subiektywnych barw – nawet jeśli na przykład to, jak czerwone rzeczy wyglądają dla ciebie, jest tym, jak zielone rzeczy wyglądają dla mnie. Nazwalibyśmy te same obiekty „czerwonymi" i „zielonymi", nawet jeśli nasze prywatne przeżycia byłyby „przeciwne" (lub po prostu inne).

Czy można jakoś stwierdzić, że tak jest? Przyjmijmy hipotezę, że czerwone rzeczy wyglądają tak samo dla ciebie i dla mnie. Czy ta hipoteza jest zarówno nie do obalenia, jak i nie do potwierdzenia? Wielu tak myślało, a niektórzy doszli do wniosku, iż z tego właśnie powodu jest to taki czy inny rodzaj nonsensu, pomimo początkowego odwołania do zdrowego rozsądku. Inni zastanawiali się, czy nie mogłaby pomóc technika, która potwierdziłaby (lub zaprzeczyła) hipotezę interpersonalnego odwróconego spektrum. Film science fiction *Burza mózgów* (śpieszę z wyjaśnieniem, że nie chodzi o ekranizację mojej książki *Brainstorms*, mimo zbieżności angielskiego tytułu – *Brainstorm*) zawierał dokładnie takie wyimaginowane urządzenie: pewien aparat neuronaukowy zakładany na głowę i przekazujący twoje przeżycia wizualne do mojego mózgu przez kabel. Z zamkniętymi oczami mogę poprawnie zrelacjonować wszystko, na co patrzysz, ale jestem zszokowany żółtym niebem, czerwoną trawą itd. Gdybyśmy mieli taką maszynę, czy taki eksperyment z jej użyciem nie potwierdziłby empirycznie hipotezy o tym, że nasze *qualia* są różne? Jednak załóżmy, że technik odłącza kabel, odwraca go o 180 stopni, ponownie podłącza i teraz mówię, że niebo jest niebieskie, trawa czerwona itd. Które z ułożeń wtyczki byłoby „poprawne"? Zaprojektowanie i zbudowanie takiego urządzenia – zakładając przez chwilę, że jest to możliwe – wymagałoby tego, by jego „wierność" była dostrojona czy skalibrowana przez normalizację sprawozdań dwóch osób badanych, więc wrócilibyśmy do punktu wyjścia. Ktoś mógłby uniknąć tej konkluzji przez dalsze wywody, jednak konsensus wśród qualofilów jest taki, że jest to przegrana sprawa; wydaje się, że istnieje powszechna umowa, iż morał z *tego* eksperymentu myślowego to brak możliwości intersubiektywnego porównania *qualiów*, nawet przy użyciu zaawansowanej techniki. Stanowi to jednak potwierdzenie słuszności szokująco „weryfikacjonistycznego" i „pozytywistycznego" punktu widzenia, że sam pomysł odwróconych *qualiów* jest nonsensem – a stąd, że sam pomysł *qualiów* jest nonsensem. Jak ujął to filozof Ludwig Wittgenstein, używając swojej słynnej analogii „żuka w pudełku",

> Rzecz w pudełku nie należy w ogóle do gry językowej; nawet nie jako *coś*, gdyż pudełko mogłoby być też puste. – Nie, przez ową rzecz w pudełku „upraszczać"; ona się znosi, czymkolwiek by była. [Wittgenstein 1953/2000, s. 144–155]

Ale co to dokładnie znaczy? Czy oznacza to, że *qualia* są prawdziwe, ale jałowe? Czy że w ogóle nie istnieją żadne *qualia*? Większości zajmującym się tym filozofom nadal wydawało się oczywiste, że *qualia* są prawdziwe, nawet jeśli różnica w *qualiach* miałaby być różnicą, której nijak nie jesteśmy w stanie wykryć. Tak miały się sprawy, nieco niepewnie, dopóki ktoś nie zastosował pozornie ulepszonej wersji tego eksperymentu myślowego: *intra*personalne odwrócone spektrum. Na pomysł wpadło najwyraźniej niezależnie kilka osób (Gert 1965; Putnam 1965; Taylor 1966; Shoemaker 1969; Lycan 1973). W tej wersji przeżycia do porównania znajdują się w jednym umyśle, więc nie potrzebujemy beznadziejnej maszyny z *Burzy mózgów*.

Pewnego dnia budzisz się i zdajesz sobie sprawę, że trawa stała się czerwona, niebo żółte itp. Nikt inny nie zauważa żadnych anomalii barw w świecie, więc problem musi tkwić w tobie. Wydaje się, że masz prawo dojść do wniosku, iż zaszła w tobie inwersja *qualiów* związanych

z barwami widzialnymi. Jak to się stało? Okazuje się, że gdy spałeś, zły neurochirurg zamienił całe okablowanie – neurony – wychodzące z wrażliwych na kolor czopków umieszczonych na twoich siatkówkach.

Brzmi nieźle. Efekt, jakiego doświadczasz, byłby zdumiewający, a może nawet przerażający. Z pewnością byłbyś w stanie wykryć, że to, jakie rzeczy wydają ci się teraz, jest różne, i nawet mielibyśmy odpowiednie naukowe wyjaśnienie dotyczące tego, dlaczego tak się stało: na przykład skupiska neuronów w korze wizualnej „zajmujące się" barwą otrzymują pobudzenia z systematycznie przesuniętych zbiorów receptorów siatkówki. Wydaje się więc, że bitwa jest w połowie wygrana: różnice *qualiów* mogą być wykrywalne, jeśli są to różnice, które pojawiły się błyskawicznie w jednej osobie[120]. Jest to jednak tylko połowa walki, gdyż wyimaginowany neurochirurgiczny żart zmienił również wszystkie twoje dyspozycje do reagowania; nie tylko *twierdzisz*, że twoje przeżycia barwne zostały postawione do góry nogami, ale twoje pozawerbalne zachowanie związane z barwami zostało również odwrócone. Podenerwowanie, które kiedyś pojawiało się przy czerwonym świetle, pojawia się teraz przy zielonym i nie masz płynności stosowania różnych systemów kodowania barwnego w twoim życiu. (Jeśli grasz w koszykówkę dla Boston Celtics, ciągle błędnie podajesz piłkę kolesiom w *czerwonych* strojach).

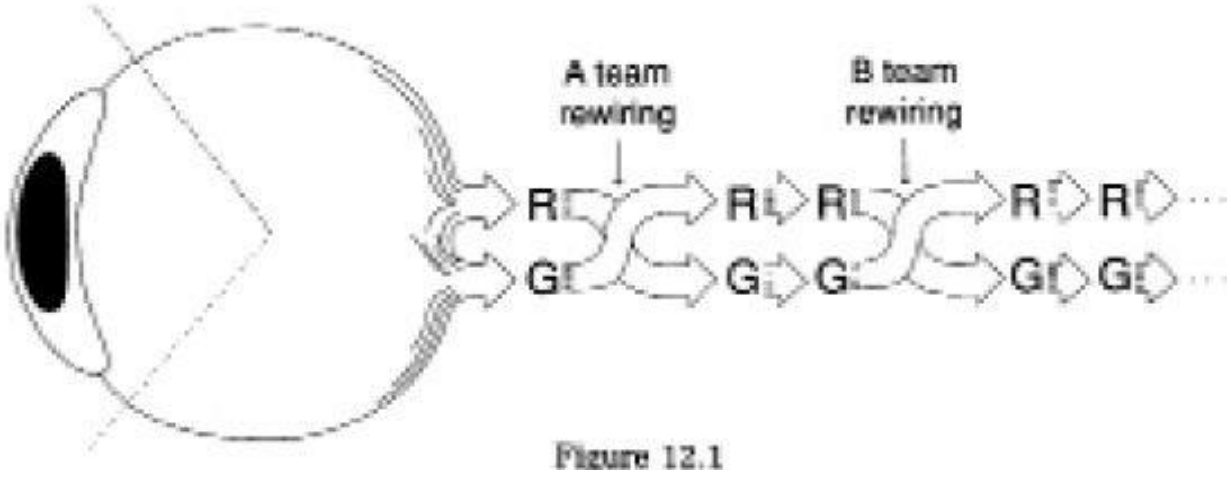


Figure 12.1

Ryc. 12.1

To, czego potrzebuje qualofil, to eksperyment myślowy pokazujący, w jaki sposób to, jak rzeczy wyglądają, może być niezależne od tych wszystkich dyspozycji do reagowania. Musimy więc dalej skomplikować historię; musimy opisać jakieś zdarzenia, które powodują cofnięcie zmiany w dyspozycjach do reagowania, jednocześnie pozostawiając nienaruszone odwrócone *qualia*. Tutaj literatura zwraca się ku jeszcze bardziej zawiłym fantazjom, gdyż nikt nawet przez chwilę nie pomyśli, że to, jak rzeczy wyglądają, będzie kiedykolwiek *rzeczywiście* odseparowane od dyspozycji podmiotu do reagowania; jest to jednak uważane za ważną *zasadniczą możliwość* przez qualofilów. Aby to pokazać, muszą opisać możliwy przypadek, jakkolwiek dziwaczny, w którym będzie oczywiste, że ta separacja rzeczywiście nastąpiła. Spójrzmy na historię, która *nie* zadziała:

Pewnej nocy, gdy śpisz, źli neurochirurdzy zmieniają wszystkie połączenia od czopków (jak wcześniej), a następnie, później, podczas tej samej nocy, przybywa kolejna grupa neurochirurgów, zespół B, i przeprowadza *uzupełniającą* zamianę nieco wyżej na nerwach optycznych.

Przywraca to wszystkie stare dyspozycje do reagowania (możemy tak założyć), ale niestety również przywraca stare *qualia*. Na przykład komórki w korze „zajmujące się" barwą będą teraz ponownie otrzymywać oryginalny sygnał dzięki szybkiemu naprawieniu szkody przez zespół B. Druga zmiana nastąpiła, jak się wydaje, zbyt wcześnie; doszło do niej *na drodze* do świadomego przeżycia. Będziemy zatem musieli opowiedzieć historię inaczej, a druga zmiana będzie musiała zachodzić później, *po tym*, jak odwrócone *qualia* dotrą do świadomości, ale *zanim* którakolwiek z odwróconych reakcji będzie mogła zostać uruchomiona. Czy jest to jednak możliwe? Nie, jeśli argumenty za modelem wielokrotnych szkiców są poprawne. Nie można zakreślić takich granic na przyczynowym „łańcuchu" od gałki ocznej przez świadomość do następującego potem zachowania, aby wszystkie reakcje na *x* następowały po nim, a świadomość *x* wystąpiła przed nim. A to dlatego, że nie jest to prosty łańcuch przyczynowy, lecz przyczynowa sieć z wieloma ścieżkami, na których wielokrotne szkice są redagowane symultanicznie i na wpół niezależnie. Historia qualofila miałaby sens, gdyby istniał teatr kartezjański, specjalne miejsce w mózgu, gdzie występuje świadome przeżycie. Gdyby takie miejsce istniało, moglibyśmy odizolować je dwoma zamianami, pozostawiając odwrócone *qualia* w teatrze, jednocześnie podtrzymując normę w dyspozycjach do reagowania. Skoro jednak nie istnieje teatr kartezjański, ten eksperyment myślowy nie ma sensu. Nie sposób spójnie opowiedzieć potrzebnej historii. Nie sposób wyodrębnić właściwości *przedstawianych* w świadomości od wielu reakcji mózgu na jego rozróżnienia, ponieważ nie istnieje tego rodzaju dodatkowy proces prezentacji.

W literaturze na temat odwróconego spektrum druga zamiana zwykle ma się dokonać nie przez operację, ale przez stopniową adaptację podmiotu do nowego trybu przeżyć. Ma to pozorny sens; ludzie potrafią zadziwiająco dobrze dostosowywać się do przedziwnych przemieszczeń zmysłów. Odbyło się wiele doświadczeń badających odwrócenie pola widzenia, w których badani nosili gogle obracające wszystko do góry nogami – przez obrót wszystkiego do normalnej pozycji! (Np. Stratton 1896; Kohler 1961; Welch 1978 dostarcza dobrego streszczenia; zob. również Cole 1990). Po kilku dniach ciągłego noszenia takich czy innych odwracających gogli (ma to znaczenie – niektóre rodzaje miały szerokie pole widzenia, a inne dawały obserwatorom rodzaj widzenia tunelowego) badani często dostosowywali się zaskakująco dobrze. W filmie Ivo Kohlera o jego eksperymencie w Innsbrucku widzimy dwóch badanych, komicznie bezradnych przy pierwszym założeniu gogli, zjeżdżających na nartach i jeżdżących na rowerze pośród samochodów, nadal z założonymi goglami, do których najwyraźniej się przyzwyczaili.

Załóżmy więc, że stopniowo dostosowujesz się do chirurgicznej inwersji twojego widzenia barwnego. (Dlaczego ktoś chciałby albo musiałby się dostosować, to inna kwestia, lecz możemy równie dobrze ustąpić tu qualofilom, aby przyśpieszyć ich upadek). Pewne adaptacje z początku byłyby wyraźnie poprzeżyciowe. Możemy założyć, że czyste niebo nadal *wydawałoby* ci się żółte, ale trzeba by je nazywać niebieskim, aby dogadać się z sąsiadami. Patrzenie na nowy przedmiot mogłoby wywołać chwilową dezorientację: „To zie... To znaczy *czerwony*!". A co z twoim podenerwowaniem przy zielonym świetle – czy nadal pojawiałoby się jako anomalia w twojej reakcji skórno-galwanicznej? Na potrzeby dyskusji qualofil musi sobie wyobrazić, jakkolwiek mało prawdopodobne by to było, że *wszystkie* twoje dyspozycje do reagowania dostosowują się do sytuacji, pozostawiając jedynie ślad nadal odwróconych *qualiów*, więc na potrzeby dyskusji przyznajmy, że najbardziej fundamentalne i wrodzone skłonności w twojej przestrzeni *qualiów* również „dostosowują się" – jest to bzdurne, ale będzie jeszcze gorzej.

Aby opowiedzieć potrzebną historię, qualofil musi założyć, że w końcu wszystkie te adaptacje stają się drugą naturą – szybką i niezbadaną. (Gdyby nie stawały się drugą naturą, pozostałyby dyspozycje do reagowania, które nadal byłyby inne, a dyskusja wymaga, aby

wszystkie zostały poprawione). Niech tak będzie. Teraz zakładając, że *wszystkie* twoje dyspozycje do reagowania zostały przywrócone, jakie są twoje intuicje co do twoich *qualiów*? Czy nadal są odwrócone, czy nie?

Wolno w tym momencie spasować, gdyż po tym, jak poproszono nas o tolerowanie tak wielu wątpliwych założeń na potrzeby dyskusji, można albo czuć pustkę – nie mieć żadnej intuicji – albo nie ufać swoim intuicjom bez względu na to, jakie one są. Jednak być może rzeczywiście wydaje się dość oczywiste, że twoje *qualia* nadal będą odwrócone. Ale dlaczego? Co w tej historii doprowadziło cię do takiego wniosku? Być może, mimo że działasz zgodnie ze wskazaniami, niewinnie dodajesz jakieś dalsze założenia niewymagane przez historię albo nie udało ci się dostrzec pewnych możliwości przez nią niewykluczonych. Uważam, iż najbardziej prawdopodobne wyjaśnienie twojej intuicji, że w tym wyimaginowanym momencie nadal pozostałyby „odwrócone *qualia*", jest takie, że dodatkowo i w sposób nieuzasadniony zakładasz, iż wszystkie te adaptacje zachodzą „po stronie poprzeżyciowej".

Mogłoby jednak być tak, że adaptacja dokonuje się w drodze wstępującej, prawda? Kiedy po raz pierwszy zakładasz mocno przyciemniane gogle, nie widzisz żadnej barwy – a przynajmniej barwy, które widzisz, są dziwne i trudne do rozróżnienia – jednak po jakimś czasie powraca zaskakująco normalne widzenie barwne. (Cole 1990 zwraca uwagę filozofów na ten efekt, który można przetestować samodzielnie z użyciem wojskowych gogli snajperskich na podczerwień). Być może, nie wiedząc o tym zaskakującym fakcie, nigdy nie przyszło ci na myśl, że *można by* dostosować się w ten sam sposób do naszej operacji. Moglibyśmy podkreślić tę możliwość w eksperymencie myślowym, dodając kilka szczegółów:

... A gdy następowała adaptacja, często dziwiło cię, że barwy przedmiotów koniec końców nie wydawały się takie dziwne, a czasem dezorientacja zwyciężała i poprawki trzeba było robić dwa razy. Gdy pytano cię o barwę nowego przedmiotu, odpowiedzią było „Jest zie..., nie, czer..., nie, jest zielony!".

Historia opowiedziana w ten sposób mogłaby sprawić, że „oczywiste" wydawałoby się dopasowanie się lub ponowne odwrócenie samych *qualiów* barw. W każdym razie możesz teraz przypuszczać, że musiało się to stać na jeden z tych sposobów. Nie mogłoby być przypadku, w którym nie było w sposób oczywisty jasne, jaki rodzaj adaptacji się pojawił! Bezkrytycznie przyjęte założenie uzasadniające *to* przekonanie jest takie, że wszystkie adaptacje mogą zostać skategoryzowane jako przed- lub poprzeżyciowe (stalinowskie lub orwellowskie). Początkowo założenie to może wydawać się niewinne, gdyż skrajne przypadki łatwo jest sklasyfikować. Gdy mózg kompensuje ruchy głowy i oka, tworząc stabilny wizualny świat „w przeżyciu", jest to z pewnością przedprzeżyciowe wymazanie, adaptacja na ścieżce do świadomości. A gdy wyobrażasz sobie peryferyjną („późną") kompensację w doborze nazw kolorów („To jest zie... to znaczy *czerwone*!"), jest to oczywiście poprzeżyciowa, jedynie behawioralna adaptacja. Jest więc chyba sensowne, że gdy *wszystkie* adaptacje zostały zakończone, to pozostawiają subiektywną barwę (barwę „w świadomości") odwróconą lub nie? Oto co byśmy powiedzieli: Dodaj zmiany na ścieżce w kierunku świadomości; jeśli liczba jest parzysta – jak po akcji zespołu B – *qualia* są znormalizowane; jeśli jest nieparzysta, *qualia* nadal są odwrócone. Nonsens. Przypomnijmy sobie krzywą neo-Laffera z rozdziału 5. Nie jest logiczną ani geometryczną koniecznością, aby istniała pojedyncza wartość rozróżnionej zmiennej, która może być wyróżniona jako *unikatowa* wartość zmiennej „w świadomości".

Możemy to zademonstrować naszą własną małą fantazją, grając według zasad qualofila. Załóżmy, że przed operacją pewien odcień niebieskiego przywoływał zwykle na myśl samochód, którym kiedyś spowodowałeś wypadek, więc była to barwa, której unikasz. Załóżmy, że po operacji z początku nie masz negatywnych reakcji na tę barwę, odbierając ją jako nieszkodliwy

i niekojarzący się z niczym żółty. Gdy jednak kończysz adaptację, ponownie unikasz rzeczy w tym odcieniu niebieskiego *i robisz to dlatego, że przypominają ci o wypadku*. (Gdyby tak nie było, byłaby to niedostosowana dyspozycja do reagowania). Ale gdy zapytamy cię, czy to dlatego, że kiedy myślisz o wypadku, pamiętasz samochód jako żółty – taki jak ten nieprzyjemny przedmiot przed tobą – czy dlatego, że gdy przypominasz go sobie, samochód jest niebieski – taki jak ten nieprzyjemny obiekt przed tobą, nie możesz odpowiedzieć. Twoje werbalne zachowanie będzie zupełnie „dostosowane"; twoja natychmiastowa, płynąca z drugiej natury odpowiedź na pytanie: „W jakim kolorze był ten samochód?" będzie „niebieski" i bez wahania powiesz, że nieprzyjemny przedmiot leżący przed tobą jest również niebieski. Czy to oznacza, że długi okres szkolenia *uległ zapomnieniu*?

Nie. Nie potrzebujemy niczego tak dramatycznego jak amnezja, aby wyjaśnić twoją niemożność odpowiedzi, bo mamy mnóstwo codziennych przypadków, w których to zjawisko się pojawia. Lubisz piwo? Wiele osób lubiących piwo przyznaje, że jest ono smakiem nabytym. Stopniowo uczymy się czerpania przyjemności ze smaku – albo z czasem dochodzimy do niej. Jakiego smaku? Smaku pierwszego łyka?

Nikt nie może lubić *tego* smaku – mógłby odpowiedzieć doświadczony piwosz. – Piwo smakuje inaczej dla doświadczonego amatora piwa. Gdyby piwo nadal smakowało mi tak, jak pierwszy jego łyk, nigdy więcej bym go nie wypił! Albo, patrząc na to z innej strony, gdyby mój pierwszy łyk piwa smakował mi w sposób, w jaki smakował mi ostatni, niedawny łyk, nie musiałbym w ogóle nabywać upodobania do piwa. Pierwszy łyk zasmakowałby mi tak samo jak ten, którym właśnie się cieszyłem.

Jeśli ten amator piwa ma rację, wówczas piwo *nie* jest nabytym smakiem. Nikt nie cieszy się tym, *jak smakuje pierwszy łyk*. Zamiast tego stopniowo zmienia się dla niego to, *jak smakuje piwo*. Inni miłośnicy piwa mogliby stwierdzić, że piwo smakuje im teraz tak samo, jak smakowało kiedyś, tylko że teraz ten smak lubią. Czy jest tu jakaś różnica? Z pewnością jest różnica w heterofenomenologii i musi ona zostać wyjaśniona. *Mogłoby* być tak, że różne przekonania wyrastają z prawdziwych różnic w umiejętnościach rozróżniania tego rodzaju: w pierwszym miłośniku piwa „trening" zmienił „kształt" przestrzeni *qualiów* związanych ze smakowaniem, a w drugim przestrzeń *qualiów* pozostaje z grubsza taka sama, ale „funkcja ewaluacyjna" tej przestrzeni została zmieniona. *Mogłoby* również być tak, że niektórzy albo i wszyscy amatorzy piwa oszukują się (jak ci, którzy niezmiennie twierdzą, że wszystkie Marilyn w wysokiej rozdzielczości naprawdę są w tle ich pola widzenia). Musimy spojrzeć ponad światy heterofenomenologiczne i przyjrzeć się rzeczywistym zdarzeniom w głowie, aby zobaczyć, czy istnieje trzymająca się prawdy (choć być może „naciągnięta") interpretacja twierdzeń miłośników piwa, a *jeśli* istnieje, to tylko dlatego, że zdecydujemy się *zredukować* „to, jak smakuje" do takiego czy innego kompleksu dyspozycji do reagowania (Dennett 1988a). Musielibyśmy „zniszczyć" *qualia* po to, aby je „ocalić".

Jeśli zatem piwosz marszczy brwi, przyjmuje śmiertelnie poważny wyraz twarzy i mówi, że to, co ma na myśli, to „*sposób, w jaki piwo teraz dla mnie smakuje*", z pewnością oszukuje siebie, jeżeli uważa, że może w ten sposób odnieść się do znanego sobie *quale*, subiektywnego stanu niezależnego od jego zmieniających się postaw związanych z reakcjami. Może mu się wydawać, że może, ale to nieprawda[121].

Dlatego też i w wyimaginowanym przypadku przypominania sobie o rozbitym samochodzie przez spojrzenie na niebieski przedmiot, samooszukiwaniem byłoby twierdzenie, że na podstawie *tego, jak ten przedmiot dla ciebie wygląda*, możesz powiedzieć, czy było to „wewnętrznie" to samo, jak wyglądał dla ciebie samochód, gdy się rozbił. To wystarczy do zniweczenia eksperymentu myślowego qualofila, gdyż celem było opisanie przypadku, w którym

byłoby *oczywiste*, że *qualia* są odwrócone, a jednocześnie dyspozycje do reagowania są znormalizowane. *Założenie*, że każdy po prostu by to wiedział, jest niemądre, a bez tego założenia nie ma argumentu, lecz jedynie pompa intuicji – historia nakłaniająca cię do zadeklarowania prawdziwości intuicji bez zaoferowania żadnej dobrej racji po temu.

Czy to niemądre, czy nie, *nadal* może wydawać się absolutnie oczywiste, że „subiektywne barwy widziane na przedmiotach" *musiałyby* być „takie czy inne". Pokazuje to siłę przyciągania teatru kartezjańskiego oddziałującego na nasze wyobraźnie. Aby pozbyć się powabu tej idei, warto dalej rozważyć porównanie z goglami odwracającymi obraz. Gdy adaptacje osób noszących te gogle stają się tak bardzo ich drugą naturą, że mogą jeździć na rowerach i na nartach, naturalnym (aczkolwiek mylącym) pytaniem może być: Czy dostosowali się, *ponownie obracając obraz przeżywanego świata*, czy *przyzwyczajając się do obrazu przeżywanego świata do góry nogami*? I cóż odpowiadają? Mówią różne rzeczy, co jest z grubsza skorelowane z tym, jak pełna była ich adaptacja. Im była pełniejsza, tym bardziej odrzucali oni pytanie jako niewłaściwe lub pozbawione odpowiedzi. Tego właśnie wymaga model wielokrotnych szkiców: skoro istnieje mnóstwo rozróżnień i reakcji rozrzuconych po mózgu, które muszą zostać dostosowane, niektóre z nich zajmujące się „odruchami" na niższym poziomie (na przykład odchylaniem się w odpowiednim kierunku, gdy coś zbliża się w twoim kierunku), a inne celowymi czynnościami będącymi w centrum uwagi, nie jest zaskakujące, że gdy adaptacje w tym patchworku się akumulują, osoby badane tracą całe przekonanie o tym, czy powinny powiedzieć „wszystko wygląda tak jak kiedyś" zamiast „wszystko nadal wygląda inaczej, ale przyzwyczajam się do tego". W pewien sposób rzeczy wyglądają dla nich tak samo (co można ocenić po ich reakcjach), w inny – wyglądają odmiennie (co można ocenić po innych reakcjach). Być może gdyby istniała jedna reprezentacja przestrzeni wizualno-motorycznej, przez którą musiałyby przechodzić wszystkie reakcje na bodźce wizualne, musiałaby *ona* być „taka albo inna", ale takiej reprezentacji nie ma. To, jak rzeczy dla nich wyglądają, składa się z wielu częściowo niezależnych nawyków do reagowania, a nie z jednego wewnętrznie ułożonego normalnie lub do góry nogami obrazu w głowie. Liczy się to, co jest między wejściem a wyjściem, a skoro jest to uzyskiwane w wielu różnych miejscach, wieloma różnymi i w dużej mierze niezależnymi środkami, po prostu nie można stwierdzić, co się „liczy" jako „moje pole widzenia jest nadal do góry nogami".

To samo dotyczy odwróconych *qualiów*. Idea, że jest to coś *dodatkowego* poza inwersją wszystkich dyspozycji do reagowania, a więc gdyby zostały zmienione ponownie, odwrócone *qualia* pozostałyby nadal odwrócone, należy jedynie do nieustępliwego mitu teatru kartezjańskiego. Mit ten pozostaje sławny dzięki rozbudowanym eksperymentom myślowym dotyczącym inwersji spektrum, jednak sława nie oznacza zasadności ani dowodów. *Jeśli* nie istnieją *qualia* poza sumą dyspozycji do reagowania, idea podtrzymywania niezmienionych *qualiów*, podczas gdy dostosowuje się dyspozycje, jest wewnętrznie sprzeczna.

5. *Qualia* „epifenomenalne"?

Istnieje kolejny eksperyment myślowy związany z przeżywaniem barw, który okazuje się nieodparcie przyciągający: stworzony przez Franka Jacksona, szeroko dyskutowany przypadek Marii, specjalistki zajmującej się barwami, która sama nigdy barw nie widziała. Jak w każdym dobrym eksperymencie myślowym, jego cel jest natychmiast ewidentny, nawet dla niewtajemniczonych. Jest on pompą intuicji zachęcającą nas do niezrozumienia swoich założeń!

Maria jest genialną badaczką, która z jakiegoś powodu jest zmuszona do badania świata z czarno-białego pokoju *przez* czarno-biały monitor telewizyjny. Specjalizuje się

w neuropsychologii widzenia i załóżmy, że zdobywa kompletne informacje fizyczne, jakie można zdobyć na temat tego, co dzieje się, gdy widzimy dojrzałe pomidory czy niebo i używamy pojęć takich jak *czerwony*, *niebieski* itp. Na przykład odkrywa, dokładnie które kombinacje długości fal z nieba pobudzają siatkówkę i dokładnie jak wytwarza to *w* centralnym układzie nerwowym napięcie strun głosowych i wyrzucenie powietrza z płuc, czego rezultatem jest wypowiedzenie zdania „Niebo jest niebieskie". [...] Co się stanie, gdy Maria zostanie wypuszczona z czarno-białego pokoju, albo gdy otrzyma kolorowy ekran? Czy *dowie* się czegoś, czy nie? Wydaje się oczywiste, że dowie się czegoś o świecie i naszym wizualnym przeżywaniu tego świata. Nie można zatem nie zauważyć, że jej wcześniejsza wiedza była niepełna. Jednak miała *wszystkie* informacje fizyczne. A zatem istnieje coś więcej niż tylko one, a fizykalizm nie jest prawdą. [Jackson 1982, s. 128]

Puenta nie mogłaby być jaśniejsza. Maria *nie* miała żadnego przeżycia barw (nie ma luster, aby mogła zobaczyć swoją twarz, jest zmuszana do noszenia czarnych rękawiczek itp.), więc w tym ważnym momencie, gdy jej porywacze w końcu pozwalają jej wyjść do świata zewnętrznego, który zna tylko z opisów (oraz czarno-białych rysunków), „wydaje się oczywiste", jak mówi Jackson, że czegoś się dowie. Możemy rzeczywiście wyraźnie ją sobie wyobrazić, patrzącą po raz pierwszy na czerwoną różę i wykrzykującą: „A więc to *tak* wygląda czerwony!". Może też wydawać się nam, że jeśli pierwszymi kolorowymi rzeczami, które zostaną jej pokazane, będą na przykład nieopisane drewniane klocki, a powie jej się tylko, że jeden z nich jest czerwony, a drugi niebieski, nie będzie miała bladego pojęcia, który jest który, dopóki jakoś nie dowie się, które nazwy barw pasują do jej nowo odkrytych przeżyć.

Tak właśnie wyobrażają sobie ten eksperyment myślowy prawie wszyscy – nie tylko niewtajemniczeni, ale nawet najbardziej przebiegli, zahartowani w walce filozofowie (Tye 1986; Lewis 1988; Loar 1990; Lycan 1990; Nemirov 1990; Harman 1990; Block 1990; van Gulick 1990). Jedynie Paul Churchland (1985, 1990) wyraził sprzeciw wobec *obrazu* tak wyraźnie wymalowanego przez eksperyment myślowy, dramatycznego odkrycia Marii. Obraz jest błędny; jeśli tak sobie wyobrażasz ten przypadek, po prostu nie postępujesz zgodnie z zawartymi w nim instrukcjami! Powód, dla którego nikt nie postępuje zgodnie ze wskazówkami, jest taki, że to, co masz sobie wyobrazić, jest tak absurdalnie przepastne, iż nie warto nawet próbować. Najważniejszym założeniem jest to, że „zdobywa kompletne informacje fizyczne". Nie jest to łatwe do wyobrażenia, więc nikt się tym nie przejmuje. Każdy wyobraża sobie tylko, że Maria wie bardzo dużo – być może, że wie wszystko, co można wiedzieć *dziś* o neurofizjologii widzenia barwnego. Ale to tylko kropla w morzu i nic dziwnego, że Maria dowiedziałaby się czegoś, gdyby *to* było wszystko, co wiedziała.

Aby odkryć tę iluzję wyobraźni, pozwólcie, że będę kontynuował historię w zaskakujący – ale zasadny – sposób:

Tak więc pewnego dnia porywacze Marii postanowili, że nadszedł czas, by zobaczyła barwy. W ramach podstępu jako pierwsze barwne przeżycie dla Marii przygotowali jasnoniebieskiego banana. Maria spojrzała nań i odrzekła: „Hej! Chcieliście mnie oszukać! Banany są żółte, a ten jest niebieski!". Jej porywacze oniemieli. Jak to zrobiła? „To proste – odpowiedziała. – Musicie pamiętać, że ja wiem *wszystko* – absolutnie wszystko – czego kiedykolwiek można by się dowiedzieć o fizycznych przyczynach i skutkach widzenia barwnego. Więc oczywiście, zanim przynieśliście banana, zapisałam sobie najdokładniej, jakie wrażenie fizyczne wywarłby na moim układzie nerwowym żółty przedmiot albo niebieski przedmiot (albo zielony przedmiot itd.). Wiedziałam więc, dokładnie jakie *myśli* bym miała (no bo w końcu »sama skłonność« do myślenia o tym czy tamtym nie jest jednym z waszych słynnych *qualiów*, prawda?). Nie byłam wcale zaskoczona moim przeżyciem niebieskiego (zaskoczyło mnie to, że

zdecydowaliście się na tak kiepski podstęp). Zdaję sobie sprawę z tego, że *jest wam trudno sobie wyobrazić*, iż mogę wiedzieć tak wiele o moich dyspozycjach do reagowania, że sposób, w jaki wpłynął na mnie niebieski, nie był żadną niespodzianką. Oczywiście, że jest wam to sobie trudno wyobrazić. Każdemu jest trudno wyobrazić sobie konsekwencje tego, że ktoś wie absolutnie wszystko, co fizyczne, o wszystkim!".

Z pewnością oszukiwałem, pomyślisz. Muszę ukrywać jakąś możliwość za zasłoną odpowiedzi Marii. Czy możesz to udowodnić? Nie chodzi mi o to, że druga część tej historii dowodzi tego, iż Maria *niczego* się nie dowiedziała, ale o to, że tradycyjny sposób wyobrażania sobie tej historii *nie dowodzi*, iż jest przeciwnie. Nie dowodzi niczego; jedynie podaje w wątpliwość intuicję, że Maria czegoś się dowiedziała („wydaje się oczywiste"), próbując cię namówić do wyobrażenia sobie czegoś innego niż to, czego wymagają przesłanki.

Jest oczywiście prawdą, że w każdej realistycznej i łatwo wyobrażalnej wersji tej historii Maria dowiedziałaby się czegoś nowego, ale w każdej realistycznej i łatwo wyobrażalnej wersji może wiedzieć wiele, lecz nie wie wszystkiego o świecie fizycznym. Zwykłe wyobrażenie sobie, że Maria wie wiele i takie pozostawienie sprawy nie jest dobrym sposobem na rozpracowanie następstw jej „zdobycia kompletnych informacji fizycznych" – nie jest nim więcej niż wyobrażenie sobie, że to, iż jest obrzydliwie bogata, jest dobrym sposobem na rozpracowanie następstw hipotezy, że wszystko należy do niej. Mogłoby nam pomóc wyobrażenie sobie stopnia mocy, jaką daje jej władza, jeśli zaczniemy od wymienienia kilku rzeczy, które z pewnością od początku wie. Wie, że czarny i biały to odcienie szarości, zna różnice między kolorem jakiegoś przedmiotu a takimi właściwościami powierzchni jak błyszczenie i matowość, a także wie wszystko o różnicach między granicami jasności a granicami barw (ogólnie rzecz biorąc, granice jasności to te widoczne na czarno-białym telewizorze). Wie również, dokładnie jaki wpływ – opisany w języku neurofizjologii – będzie miała każda konkretna barwa na jej układ nerwowy. A zatem pozostaje jedynie, aby wymyśliła, jak identyfikować te efekty neurofizjologiczne „od wewnątrz". Być może łatwo ci wyobrazić sobie, że robi w tym kierunku *małe* postępy – na przykład wymyślając sprytne sposoby, które pomogą jej stwierdzić, że jakaś barwa, nieważne jaka, *nie* jest żółta, *nie* jest czerwona. Jak? Zauważając jakąś wyraźną i charakterystyczną reakcję, którą jej mózg miałby tylko na żółć albo tylko na czerwień. Jeśli jednak pozwolisz jej nawet na tak niewielkie spojrzenie na przestrzeń barw, musisz dojść do wniosku, że może uzyskać ostatecznie zaawansowaną wiedzę, gdyż nie zna jedynie *wyraźnych* reakcji, zna je wszystkie.

Przypomnij sobie kartonik po galaretce Juliusa i Ethel Rosenbergów, który zamienili w detektor *M*. A teraz wyobraź sobie ich zdziwienie, gdyby pojawił się oszust z „pasującym" kawałkiem, który nie był oryginałem. „Niemożliwe!" – krzyczą. „Nie niemożliwe – mówi oszust – jedynie trudne. Miałem *kompletne informacje fizyczne* konieczne do rekonstrukcji detektora *M* oraz do zrobienia kolejnego przedmiotu z właściwością kształtu *M*". Maria miała wystarczającą ilość informacji (w oryginalnym przypadku, odpowiednio wyobrażanym), aby zorientować się, dokładnie czym były detektory czerwieni i niebieskości, więc mogła zidentyfikować je zawczasu. Nie jest to typowy sposób uczenia się barw, ale Maria nie jest typowym człowiekiem.

Wiem, że nie usatysfakcjonuje to wielu filozoficznych fanów Marii oraz że pozostało jeszcze sporo do wyjaśnienia, ale – i oto sedno tego wszystkiego – rzeczywiste udowadnianie musi przebiegać na terenie bardzo dalekim od przykładu Jacksona, który to przykład wywołuje klasyczny syndrom filozofa: pomylenie porażki wyobraźni z wglądem w konieczność. Niektórzy filozofowie zajmujący się przypadkiem Marii mogą nie dbać o to, że wyobrazili go sobie źle, bo po prostu użyli go jako trampoliny do dyskusji rzucających światło na przeróżne niezależnie interesujące i istotne kwestie. Nie będę się tymi kwestiami tu zajmować, ponieważ jestem

zainteresowany bezpośrednią analizą wniosku wyciągniętego z tego przykładu przez samego Jacksona: przeżycia wizualne mają *qualia*, które są „epifenomenalne".

Pojęcie „epifenomen" jest dziś w powszechnym użyciu zarówno wśród filozofów, jak i psychologów (oraz innych kognitywistów). Używane jest przy założeniu, że jego znaczenie jest powszechnie znane i niebudzące zastrzeżeń, gdy tak naprawdę filozofowie i kognitywiści stosują je *zupełnie* inaczej – dziwny fakt, o tyle dziwniejszy, że mimo wielokrotnych prób zmiany tej problematycznej sytuacji nikogo chyba ona nie obchodzi. Skoro „epifenomenalizm" zwykle wydaje się ostatnim bezpiecznym schronieniem *qualiów* oraz skoro to pozorne schronienie całkowicie wynika z zamieszania wywołanego właśnie tymi niespójnymi znaczeniami, muszę zepchnąć jego obrońców do pozycji obronnych.

Według *Shorter Oxford English Dictionary* pojęcie „epifenomen" po raz pierwszy pojawia się w 1706 roku jako pojęcie z dziedziny patologii, „wtórny objaw lub symptom". Biolog ewolucyjny Thomas Huxley (1874) prawdopodobnie był autorem, który rozszerzył znaczenie tego słowa do jego obecnego użycia w psychologii, gdzie oznacza *niefunkcjonalną* właściwość lub produkt uboczny. Huxley używał tego terminu w swoich rozważaniach o ewolucji świadomości i w twierdzeniu, że właściwości epifenomenalne (jak „gwizd maszyny parowej") nie mogły zostać wyjaśnione przez dobór naturalny.

Oto wyraźny przykład użycia tego słowa:

Dlaczego ludzie głęboko myślący przygryzają wargę i stukają palcami? Czy te czynności to epifenomeny towarzyszące głównemu procesowi czucia i myślenia, czy mogą same w sobie być integralną częścią tych procesów? [Zajonc i Markus 1984, s. 74]

Zwróćmy uwagę, że autorzy chcą zapewnić, iż te czynności, choć łatwo zauważalne, nie odgrywają żadnej istotnej roli, czy też wyznaczonej roli, w procesie czucia i myślenia; są niefunkcjonalne. Tak samo szum komputera jest epifenomenalny, tak jak twój cień, gdy robisz sobie filiżankę herbaty. Epifenomeny to zwykłe produkty uboczne, ale jako takie są produktami z wieloma efektami w świecie; stukanie palcami wywołuje słyszalny dźwięk, a twój cień ma wpływ na fotografię, nie wspominając o delikatnym chłodzeniu powierzchni, na które pada.

Standardowe znaczenie filozoficzne jest inne: „x jest epifenomenalne" oznacza „x jest efektem, ale samo w sobie nie ma żadnych efektów w świecie fizycznym". (Broad 1925, s. 118, podaje definicję rozpoczynającą, a przynajmniej utwierdzającą, użycie filozoficzne). Czy te znaczenia rzeczywiście są tak różne? Tak, są tak różne, jak znaczenie wyrazów „*morderstwo*" i „*śmierć*". Znaczenie filozoficzne jest mocniejsze: wszystko, co nie ma żadnego efektu w świecie fizycznym, z pewnością nie ma wpływu na żadną funkcję, ale nie jest tak w psychologii, jak wyraźnie widać w przykładzie Zajonca i Markusa.

Znaczenie filozoficzne jest tak naprawdę za mocne; daje w efekcie zupełnie bezużyteczne pojęcie (Harman 1990; Fox 1989). Skoro x nie ma efektów fizycznych (według tej definicji), żadne narzędzie nie może wykryć obecności x, bezpośrednio i pośrednio; procesy w świecie nie są w żaden sposób zmieniane przez obecność lub nieobecność x. Jak więc kiedykolwiek mogła się pojawić empiryczna racja uzasadniająca uznanie istnienia x? Załóżmy na przykład, iż Otto uważa, że właśnie on ma epifenomenalne *qualia*. Dlaczego to mówi? Nie dlatego, że mają na niego jakiś wpływ, jakoś go prowadząc czy ostrzegając, gdy to deklaruje. Z samej definicji epifenomenów (w sensie filozoficznym) gorące wyznania Ottona, że ma epifenomeny, *nie mogą* być dla niego ani dla nikogo innego dowodem tego, że rzeczywiście je ma, bo mówiłby dokładnie to samo, nawet gdyby ich nie miał. Ale może Otto ma jakieś „wewnętrzne" świadectwo?

Jest tu mały kruczek, choć wcale nie atrakcyjny. Pamiętaj, że epifenomeny są definiowane jako niemające żadnego wpływu na świat *fizyczny*. Jeśli Otto chce przyjąć absolutny

dualizm, może twierdzić, że jego epifenomenalne *qualia* nie mają wpływu na świat fizyczny, lecz mają wpływ na jego (niefizyczny) świat umysłu (Broad 1925 pozbawił nas tego kruczka przez swoją definicję, ale nie zaszkodzi go rozważyć). Na przykład *powodują niektóre z jego (niefizycznych) przekonań*, jak przekonanie, że ma epifenomenalne *qualia*. Jest to jednak tylko tymczasowa ucieczka od opresji. Otto, nie chcąc przeczyć sobie samemu, musiałby uznać swoje przekonania za niemające wpływu na świat fizyczny. Gdyby nagle utracił swoje epifenomenalne *qualia*, nie wierzyłby już w to, że je ma, ale nadal *mówiłby*, że je ma. Po prostu sam nie wierzyłby w to, co mówi! (Ani nie mógłby powiedzieć, że nie wierzy w to, co mówi, ani zrobić nic, co pokazałoby, że nie wierzy już w to, co mówi). A zatem jedyny sposób, w jaki Otto mógłby „uzasadnić" swoje przekonanie o epifenomenach, to wycofanie się do solipsystycznego świata, gdzie znajduje się tylko on, jego przekonania i jego *qualia*, odcięci od wszelkich efektów w świecie zewnętrznym. Nie jest to „bezpieczny" sposób na bycie materialistą i jednoczesne posiadanie *qualiów*, ale jest najlepszy, aby wesprzeć najbardziej radykalny solipsyzm, odcinając twój umysł – twoje przekonania i doświadczenia – od jakiejkolwiek relacji ze światem materialnym.

Jeśli *qualia* są epifenomenalne w standardowym sensie filozoficznym, to ich występowanie nie może wyjaśnić tego, jak cokolwiek się zdarza (w świecie materialnym), gdyż z definicji wszystko zachodziłoby dokładnie tak samo bez nich. Nie może więc być żadnego empirycznego powodu, by wierzyć w epifenomeny. Czy może być inna racja, aby uznać ich istnienie? Jaka racja? Prawdopodobnie racja *a priori*. Ale jaka? Nikt nigdy jej nie wskazał – nie słyszałem o dobrej, złej ani obojętnej. Chcącym zaprotestować i powiedzieć, że jestem w kwestii tych epifenomenów „weryfikacjonistą", odpowiadam: Czy nie każdy jest weryfikacjonistą w kwestii *takiego* rodzaju twierdzenia? Spójrzmy na przykład na hipotezę, że w każdym cylindrze wewnętrznego silnika spalinowego znajduje się czternaście epifenomenalnych gremlinów. Gremliny te nie mają masy, energii, żadnych właściwości fizycznych; nie sprawiają, że silnik pracuje bardziej lub mniej płynnie, szybciej lub wolniej. Nie ma i *nie może być* żadnego dowodu empirycznego na ich istnienie ani żadnego empirycznego sposobu na odróżnienie tej hipotezy od jej rywalek: w silniku jest dwanaście, trzynaście albo piętnaście... gremlinów. Na jakiej zasadzie bronimy całkowitego odrzucenia takiego nonsensu? Na zasadzie weryfikacjonistycznej czy po prostu zdrowego rozsądku?

Ale przecież jest różnica! – mówi Otto. – Nie ma niezależnej racji, by brać hipotezę z gremlinami na poważnie. Po prostu wymyśliłeś ją chwilę temu. *Qualia* natomiast istnieją od dłuższego czasu, odgrywając ważną rolę w naszym schemacie pojęciowym.

A co jeśli jacyś nieoświeceni ludzie myśleli od pokoleń, że to gremliny wprawiają w ruch ich samochody, ale zostali zepchnięci przez naukę na margines, lecz nadal wierzą w gremliny i w to, że są one epifenomenalne? Czy jest z naszej strony błędem od razu odrzucać ich „hipotezę"? Na jakiejkolwiek zasadzie się nie oprzemy, odrzucając taki nonsens, wystarczy, aby odrzucić doktrynę, że *qualia* są epifenomenalne w sensie filozoficznym. Nie są to poglądy, które zasługują na poważną dyskusję.

Trudno uwierzyć, że filozofowie identyfikujący ostatnio swój pogląd jako epifenomenalizm mogą popełniać tak nieszczęsny błąd. Czy może twierdzą, że *qualia* są epifenomenalne w sensie huxleyowskim? W tym wypadku *qualia są* fizycznymi skutkami i *mają* fizyczne skutki; ale nie są funkcjonalne. Każdy materialista powinien z radością przyznać, że ta hipoteza jest prawdziwa – jeśli na przykład utożsamiamy *qualia* z dyspozycjami do reagowania. Jak zauważyliśmy w rozważaniach o sprawianiu radości, nawet jeśli jakieś wybrzuszenia czy uprzedzenia w naszych przestrzeniach *qualiów* są funkcjonalne – albo kiedyś były funkcjonalne – inne są jedynie zwykłym przypadkiem. Dlaczego nie lubię brokułów? Prawdopodobnie bez

powodu; moja negatywna dyspozycja do reagowania jest czysto epifenomenalna, jest produktem ubocznym mojej budowy, nie ma żadnego znaczenia. Nie ma funkcji, ale ma mnóstwo efektów. W każdym zaprojektowanym systemie niektóre właściwości są kluczowe, podczas gdy inne mniej więcej podlegają improwizowanej korekcie. Wszystko musi być takie albo inne, ale często nie ma znaczenia, jakie jest. Skrzynia biegów w samochodzie musi mieć odpowiednią długość i siłę, ale to, czy jest w przekroju okrągła, kwadratowa czy owalna, jest właściwością epifenomenalną w sensie huxleyowskim. W systemach CAD Dla Niewidomych, o których mówiliśmy w rozdziale 10, konkretny schemat kodowania kolorów numerami był epifenomenalny. Moglibyśmy go „odwrócić" (korzystając z liczb ujemnych albo mnożąc wszystkie wartości przez jakąś liczbę), nie wprowadzając żadnej *funkcjonalniej* różnicy dla jego sprawności w przetwarzaniu informacji. Takie odwrócenie mogłoby być niewykrywalne w normalnym badaniu i mogłoby być niewykrywalne *przez system*, ale nie byłoby epifenomenalne w sensie filozoficznym. Na przykład w rejestrach pamięci przechowujących liczby byłoby sporo maleńkich różnic w napięciu.

Gdy myślimy o wszystkich właściwościach naszych układów nerwowych umożliwiających nam widzenie, słyszenie, zapach, smak i dotyk, możemy je z grubsza podzielić na właściwości odgrywające naprawdę kluczowe role w pośredniczeniu w przetwarzaniu informacji oraz na właściwości epifenomenalne, które mniej więcej podlegają improwizowanej korekcie, jak system kodowania kolorów w systemach CAD Dla Niewidomych. Gdy filozof przypuszcza, że *qualia* są epifenomenalnymi właściwościami stanów mózgu, może to oznaczać, że *qualia* mogłyby się okazać lokalnymi wariacjami w cieple generowanym przez metabolizm neuronalny. Chyba nie to mają na myśli epifenomenaliści, prawda? A jeśli tak, to *qualia* jako epifenomeny nie są wyzwaniem dla materializmu.

Nadeszła chwila, aby uczciwie przerzucić ciężar dowodu na tych, którzy upierają się przy używaniu tego pojęcia. Jego filozoficzny sens jest po prostu niedorzeczny; sens huxleyowski jest stosunkowo jasny i nieproblematyczny – ale też oderwany od argumentów filozoficznych. Żaden inny sens tego terminu nie jest w użyciu. Jeśli zatem ktoś twierdzi, że podtrzymuje jakiś rodzaj epifenomenalizmu, postaraj się zachować uprzejmość, ale zapytaj: O czym ty mówisz?

Przy okazji zwróćmy uwagę, że ta ekwiwokacja między dwoma sensami terminu „epifenomenalny" zakaża również dyskusję o zombi. Zombi filozofa, jak z pewnością pamiętasz, pod względem zachowania jest nieodróżnialny od normalnych istot ludzkich, ale nie jest świadomy. Bycie zombi nie przypomina niczego; po prostu takie się obserwatorom wydaje (łącznie z samym sobą, jak widzieliśmy w poprzednim rozdziale). Można to zinterpretować delikatnie lub zdecydowanie, w zależności od tego, jak potraktujemy nieodróżnialność dla obserwatora. Jeśli zadeklarowalibyśmy, że *zasadniczo* zombi jest nieodróżnialny od zwykłej osoby, wówczas mówilibyśmy, że prawdziwa świadomość jest epifenomenalna *w sensie absurdalnym*. Byłoby to niemądre. Zamiast tego moglibyśmy powiedzieć, że świadomość mogłaby być epifenomenalna w sensie huxleyowskim: chociaż istnieje sposób na odróżnienie zombi od zwykłych ludzi (kto wie, może zombi mają zielone mózgi), to różnica ta nie ukazuje się obserwatorom jako funkcjonalna. Tak samo ludzkie ciała z zielonymi mózgami nie skrywają obserwatorów, ale ludzkie ciała tak. Zgodnie z tą hipotezą bylibyśmy zasadniczo w stanie odróżnić ciała zamieszkałe od niezamieszkałych, sprawdzając barwę ich mózgów. To również jest niemądre, niebezpiecznie niemądre, gdyż jest echem tego samego rodzaju całkowicie bezpodstawnych uprzedzeń, które odmówiły ludziom pełni człowieczeństwa na podstawie koloru ich skóry. Czas już, aby rozpoznać ideę możliwości istnienia zombi jako to, czym naprawdę jest: nie poważnym pomysłem filozoficznym, ale niedorzecznym i haniebnym reliktem dawnych uprzedzeń. Może kobiety nie są naprawdę świadome! Może Żydzi! Cóż za szkodliwy nonsens.

Jak mówi Shylock z *Kupca weneckiego*, słusznie zwracając naszą uwagę na „zaledwie behawioralne" kryteria:

Czy Żyd nie ma oczu, czy Żyd nie ma rąk, narządów,ciała, zmysłów, uczuć i nadziei, skoro żywiony jest tą samą strawą, kaleczony tym samym orężem, podległy tym samym chorobom, leczony podobnymi sposobami, rozgrzewany i chłodzony przez to samo lato i zimę jak Chrześcijanin? – Jeśli kłujecie nas, czy nie krwawimy? Jeśli łaskoczecie nas, czy nie śmiejemy się? Jeśli trujecie nas, czy nie giniemy?

[przeł. Maciej Słomczyński]

Istnieje inne podejście do możliwości istnienia zombi i myślę, że pod pewnymi względami jest bardziej satysfakcjonujące. Czy zombi są możliwe? Nie tylko są możliwe, są rzeczywiste. Wszyscy jesteśmy zombi[122]. Nikt nie jest świadomy – nie w systematycznie tajemniczy sposób wspierający takie doktryny jak epifenomenalizm! Nie mogę udowodnić, że nie istnieje żaden taki rodzaj świadomości. Nie mogę też udowodnić, że nie istnieją gremliny. Najlepsze, co mogę zrobić, to pokazać, że nie ma żadnych porządnych powodów, aby weń wierzyć.

6. Wracając na mój fotel

W drugiej sekcji rozdziału 2 przedstawiłem zadanie wyjaśnienia świadomości, przypominając sobie epizod z mojego własnego świadomego przeżycia, gdy zacząłem bujać się w fotelu, patrząc przez okno na piękny wiosenny dzień. Powróćmy do tego fragmentu i zobaczmy, jak rozwinięta przeze mnie teoria daje sobie z nim radę. Oto tekst:

Była wczesna wiosna, zielono-złote światło słoneczne wpadało przez okno, a tysiące gałązek klonu rosnącego na podwórku było wciąż widocznych spod mgiełki zielonych pączków, tworząc elegancki deseń o wspaniałej strukturze. Szyba w oknie jest ze starego szkła i jest na niej ledwo zauważalna rysa, więc gdy bujałem się w fotelu, ta niedoskonałość spowodowała falę skoordynowanych ugięć, które poruszały się tam i z powrotem poprzez deltę gałęzi; regularny ruch, nałożony z niezwykłą wyrazistością na bardziej chaotyczne migotanie gałązek na wietrze.

Wtedy zdałem sobie sprawę, że ten wizualny metronom w gałęziach drzewa porusza się w rytmie *concerto grosso* Vivaldiego, którego słuchałem w tle. [...] Moje świadome myślenie, a szczególnie radość, jaką czułem z połączenia światła słonecznego, słonecznych skrzypiec Vivaldiego, poruszających się gałęzi – oraz przyjemność, jaką dawało mi po prostu myślenie o tym – jak to *wszystko* mogłoby być tylko czymś materialnym zachodzącym w moim mózgu? Jak jakakolwiek kombinacja elektrochemicznych zdarzeń mogłaby składać się na to, jak setki gałązek poddały się muzyce? W jaki sposób jakieś zdarzenie z zakresu przetwarzanie informacji w moim mózgu mogłoby *być* delikatnym ciepłem światła słonecznego opadającego na mnie? [...] Zdaje się to niemożliwe.

Zachęcałem nas wszystkich do stania się heterofenomenologami, więc nie mogę sam się z tego zwolnić i powinienem być zadowolony z bycia osobą badaną tak samo jak z bycia heterofenomenologiem, a zatem: zastosuję moją własną teorię do siebie. Nasze zadanie, jako heterofenomenologów, polega na przyjrzeniu się tekstowi, jego interpretacji, a następnie odniesieniu przedmiotów będących wynikiem heterofenomenologicznego świata Dennetta do zdarzeń odbywających się w tym czasie w mózgu Dennetta.

Tekst powstał kilka tygodni czy miesięcy po zdarzeniach, o których jest w nim mowa, więc możemy być pewni, że jest streszczeniem, nie tylko z powodu celowych skrótów redakcyjnych autora, ale również przez nieubłagany, streszczający proces przemijającej pamięci.

Stwierdziliśmy wcześniej – gdyby autor chwycił za dyktafon, kiedy siedział w fotelu i stworzył na miejscu tekst – że byłby on trochę inny. Nie tylko bogatszy w szczegóły i bardziej nieuporządkowany, ale oczywiście także przekształcony i przekierowany przez reakcje samego autora na proces tworzenia tekstu – słuchanie rzeczywistych dźwięków jego własnych słów, a nie samej muzyki. Jak wie każdy wykładowca, mówienie na głos często ujawnia konsekwencje (a szczególnie problemy) przekazywanej wiadomości, które umykają, gdy pogrążamy się w cichym monologu.

Tekst w takiej formie pokazuje jedynie porcję (bez wątpienia wyidealizowaną) treści świadomości autora. Musimy jednak być uważni, aby nie zakładać, że „fragmenty nieujęte" w tekście były „w rzeczywistości obecne" w czymś, co moglibyśmy nazwać „strumieniem świadomości autora". Nie możemy popełnić błędu i założyć, że istniały fakty – nie do odzyskania, ale prawdziwe fakty – dotyczące tego, które treści były w tamtym momencie świadome, a które nie. A szczególnie nie powinniśmy zakładać, że gdy autor spojrzał za okno, „przyjął to wszystko" jednym, wspaniałym łykiem umysłowym – nawet jeśli tak portretuje to ten tekst. Wydawało mu się, według tekstu, jakby jego umysł – jego pole widzenia – było wypełnione skomplikowanymi detalami złoto-zielonych pączków i migoczących gałązek, ale choć tak właśnie mu się to wydawało, była to iluzja. Żadne takie „przedstawienie" nigdy nie przyszło mu do głowy; pozostało one w świecie zewnętrznym, gdzie nie musiało być *reprezentowane*, a po prostu mogło *być*. Gdy zachwycamy się, w tych momentach uniesionej samoświadomości, wspaniałym bogactwem naszego świadomego przeżywania, to jest ono tak naprawdę bogactwem świata zewnętrznego, w całej jego olśniewającej szczegółowości. Nie „wchodzi" do naszych świadomych umysłów, lecz po prostu jest dla nich dostępne.

A co z falą skoordynowanych ugięć, które poruszały gałązkami? Gałęzie na drzewie na zewnątrz z pewnością nie poruszały się, ponieważ ruch ten był spowodowany rysą na szybie, ale nie oznacza to, że cały ten ruch musiał wystąpić w umyśle czy mózgu autora, a jedynie, że odbywał się po wewnętrznej stronie szyby, która była jego przyczyną. Gdyby ktoś sfilmował zmieniające się obrazy na siatkówkach autora, znalazłby tam ów ruch, jak w filmie, choć niewątpliwie właśnie tam cały ruch się zatrzymał; to, co działo się po wewnętrznej stronie jego siatkówek, było jedynie jego rozpoznaniem, że, jak mówi w tekście, była tam fala skoordynowanych ugięć, których doświadczał. Zobaczył ten ruch, zobaczył, w jakim stopniu się odbywał, w taki sam sposób, jak widzi się wszystkie obrazy Marilyn na tapecie. Jego siatkówkom dostarczono tę ciągłą dozę ruchu, więc gdyby chciał zasmakować ich bardziej, *pojawiłoby się* więcej szczegółów w wielokrotnych szkicach, których jedyną pozostałością jest tekst.

Było wiele innych szczegółów, na których mógł się skupić autor, ale tego nie zrobił. Istnieje wiele prawdziwych faktów, nie do odzyskania, dotyczących tego, które z tych detali zostały rozróżnione, gdzie i kiedy, przez różne systemy w jego mózgu, lecz suma tych faktów nie daje odpowiedzi na pytania takie jak: których z nich był na pewno i naprawdę świadomy (ale o nich zapomniał do czasu, gdy pisał ten tekst) oraz które były z pewnością i naprawdę na „drugim planie" jego świadomości (choć wówczas się nimi nie zajął). Nasza tendencja do zakładania, że *musi* istnieć w tych kwestiach jakiś fakt, który stanowi odpowiedź na takie pytania, jest jak naiwne założenie czytelnika, że *musi* istnieć odpowiedź na pytania typu: Czy Sherlock Holmes zjadł jajka na śniadanie w dniu, w którym poznał go doktor Watson? Conan Doyle *mógł* umieścić ten szczegół w tekście, ale tego nie zrobił, a skoro tego nie zrobił, po prostu nie istnieją żadne fakty rozstrzygające, czy te jajka należą *do fikcyjnego świata Sherlocka Holmesa*. Nawet jeśli Conan Doyle *pomyślał* o Holmesie jedzącym jajka tamtego ranka, nawet jeśli w początkowej wersji tekstu Holmes *jest na papierze przedstawiony* jako jedzący jajka

o poranku, po prostu nie ma faktu dotyczącego tego, czy w fikcyjnym świecie Sherlocka Holmesa, świecie powołanym z opublikowanego tekstu, do którego mamy dostęp, zjadł on jajka na śniadanie.

Tekst, który dostaliśmy od Dennetta, nie był „napisany w jego mózgu" między chwilą na bujanym fotelu a momentem, gdy autor zapisał go w edytorze tekstu. Koncentracja, którą osiągnął podczas bujania, oraz towarzyszące jej badanie szczegółów, które przykuły jego uwagę, miały wpływ na dość bezpieczne ugruntowanie treści tych szczegółów „w pamięci", jednak ten wpływ nie powinien być postrzegany jako przechowywanie obrazu (czy zdania) ani żadnej innej wyraźnej reprezentacji. Powinien być raczej traktowany jedynie jako sprawiający, że jakaś powtarzalność aktywności uprawdopodobnia ją bardziej i że możemy zakładać, iż takie prawdopodobne zdarzenie jest tym, co nastąpiło w momencie pisania, czyli prowadzenia demonów słów w jego mózgu w koalicje, które po raz pierwszy wytworzyły ciąg zdań. Częścią tego, co zdarzyło się wcześniej, na bujanym fotelu, bez wątpienia było zwerbowanie rzeczywistych, angielskich słów i fraz, a ta wcześniejsza współpraca między treściami bez słów a słowami niewątpliwie ułatwiła odtworzenie części właśnie tych angielskich sformułowań, gdy doszło do pisania.

Powróćmy do heterofenomenologicznego świata w tym tekście. Co z radością, o której pisze autor? „[...] połączenie światła słonecznego, słonecznych skrzypiec Vivaldiego, poruszających się gałęzi – oraz przyjemność, jaką dawało mi zwykłe myślenie o tym. [...]" Nie może to zostać wyjaśnione przez wywołanie wewnętrznie przyjemnych *qualiów* wzroku, słuchu i zwykłej myśli. Idea takich *qualiów* jedynie odwraca naszą uwagę od wszystkich możliwych ścieżek wyjaśnień i przykuwa ją w sposób, w jaki robi to kiwający palec przed oczami niemowlęcia, co sprawia, że w odrętwieniu gapimy się na „wewnętrzny obiekt", zamiast poszukiwać opisu leżących u podstaw mechanizmów *oraz* wyjaśnienia (ostatecznego i ewolucyjnego) tego, dlaczego owe mechanizmy robią to, co robią.

Przyjemność autora jest łatwa do wyjaśnienia poprzez fakt, że całe doświadczenie wizualne składa się z czynności obwodów neuronalnych, która to właśnie czynność jest dla nas wewnętrznie przyjemna nie tylko dlatego, że po prostu lubimy być poinformowani, ale również dlatego, że lubimy te konkretne sposoby, jakimi się nas informuje. Fakt, że wygląd wiosennych pąków pokrytych cętkami słońca jest czymś, co lubią ludzie, nie jest niespodzianką. To, że niektórzy ludzie lubią także patrzeć przez mikroskop na bakterie, a inni lubią oglądać zdjęcia z katastrof lotniczych, jest dziwniejsze, ale sublimacje i perwersje pragnień wyrastają z tego samego źródła w budowie naszych układów nerwowych.

Autor zastanawia się dalej, jak to możliwe, że „wszystko to mogłoby być jedynie kombinacją elektrochemicznych zdarzeń w moim mózgu". Jak jasno wskazują jego rozważania, tak się nie wydaje. A przynajmniej był moment, w którym stwierdził, że nie wydawało mu się, iż wszystko jest jedynie kombinacją elektrochemicznych zdarzeń w mózgu. Jednak nasze późniejsze rozdziały stanowią ripostę: A co by ci się wydawało, gdyby wszystko było jedynie kombinacją elektrochemicznych zdarzeń w mózgu?[123] Czy nie mamy już podstaw, by wywnioskować, że organizacja naszego mózgu prowadzi właśnie do takiego rodzaju heterofenomenologicznego świata, jakiego można by się spodziewać? Dlaczego takie kombinacje elektrochemicznych zdarzeń w mózgu nie miałyby mieć dokładnie takich efektów, które mieliśmy wyjaśnić?

(Mówi autor:) Pozostaje jednak jeszcze jedna zagadka. Skąd *ja* wiem to wszystko? Jak to jest, że *ja* mogę opowiedzieć wam wszystko o tym, co działo się w mojej głowie? Wyjaśnienie tej zagadki jest proste: *Ponieważ tym jestem*. Ponieważ ten, który wie, oraz ten, który relacjonuje na takich zasadach, to ja. *Moje* istnienie jest wyjaśnione przez fakt, że w tym ciele są takie

możliwości.

Ten pomysł jaźni jako środka narracyjnej ciężkości jest tym, co w końcu jesteśmy gotowi zbadać. Z pewnością nadszedł już czas. Wyobraź sobie moje mieszane uczucia, gdy odkryłem, że zanim udało mi się opublikować moją wersję tej idei[124], została już ona satyrycznie pokazana w powieści Davida Lodge'a *Fajna robota* (1988/1995). Jest to najwyraźniej gorący temat wśród dekonstrukcjonistów:

Według Robyn (czy też dokładniej, myślicieli, którzy wywarli istotny wpływ na jej poglądy) nie istnieje coś takiego jak „jaźń", na której opierają się zarówno kapitalizm, jak i klasyczne powieści. Nie ma więc niepowtarzalnej, wyjątkowej duszy czy esencji tworzącej osobowość jednostki, lecz tylko podmiot uwikłany w nieskończoną liczbę sieci dyskursów – dyskurs władzy, seksu, rodziny, nauki, religii, poezji i tak dalej. Na podobnej zasadzie nie można mówić o „autorze", czyli kimś, kto tworzy powieść *ab nihilo* [...]. Według znanego powiedzenia Jacques'a Derridy [...] „*il n'y a pas de hors-texte*" – nie ma nic poza tekstem. Nie istnieją źródła, lecz tylko wytwórstwo, które ma na celu stworzenie nas samych w języku. Robyn uważa, że nie „*jesteś tym, co zjesz",* ale „*jesteś tym, co mówisz*", a raczej „*jesteś tym, co cię mówi*". Zapytana, jak określić podobne stanowisko, nazwałaby je zapewne materializmem semiotycznym (Lodge 1988/1995, s. 29)[125].

Materializm semiotyczny? Czy *ja* muszę to tak nazywać? Pomijając aluzje do kapitalizmu i klasycznej powieści, do których nie zamierzam nawiązywać, ten żartobliwy fragment jest świetną parodią poglądu, który zaraz zaprezentuję. (Jak każda parodia, jest przesadny; nie powiedziałbym, że nie ma *nic* poza tekstem. *Są* na przykład regały, budynki, ciała, bakterie...)

Robyn i ja myślimy podobnie – i oczywiście *oboje* jesteśmy, jak sami to stwierdzamy, rodzajem fikcyjnych postaci, choć nieco innego rodzaju.

Rozdział 13

Realność jaźni

Przypuściwszy zaś, że istnieje maszyna, której budowa pozwala, aby myślała, czuła, miewała postrzeżenia, będzie można pomyśleć ją, z zachowaniem tych samych proporcji, tak powiększoną, aby można do niej wejść jak do młyna. Założywszy to, odnaleźlibyśmy wewnątrz przy zwiedzaniu jej tylko części, które popychają się wzajemnie, nigdy jednak nic, co tłumaczyłoby postrzeżenie.

Gottfried Wilhelm Leibniz, 1646–1716,
Monadologia (1714/1969) [przeł. Stanisław Cichowicz]

Co do mnie, to gdy wnikam najbardziej intymnie w to, co nazywam moim ja, to zawsze natykam się na jakąś poszczególną percepcję tę czy inną, ciepła czy chłodu, światła czy cienia, miłości czy nienawiści, przykrości czy przyjemności. Nie mogę nigdy uchwycić mego ja bez jakiejś percepcji i nie mogę nigdy postrzegać nic innego niż percepcję. [...] Jeżeli ktoś, zastanowiwszy się poważnie i bez uprzedzeń, myśli, iż ma inne pojęcie swego ja, to, muszę wyznać, nie mogę dłużej z nim rozumować. Wszystko, co mu mogę przyznać, to to, że on może mieć słuszność równie dobrze, jak ja i że różnimy się w sposób zasadniczy w tej sprawie. Być może jest to w jego mocy postrzegać jakąś rzecz prostą i istniejącą nieprzerwanie, którą on nazywa swoim ja, choć ja jestem pewien, że we mnie nie ma takiej rzeczy.

David Hume, 1793 [przeł. Czesław Znamierowski]

Od zarania nowoczesnej nauki w XVII wieku niemal jednogłośnie twierdzi się, że jaźń, czymkolwiek jest, byłaby niewidoczna pod mikroskopem oraz niewidoczna dla introspekcji. Dla niektórych oznacza to, że jaźń jest niefizyczną duszą, duchem w maszynie. Inni uważają, że jaźń nie jest tak naprawdę niczym, jest wymysłem metafizycznie rozpalonych wyobraźni. A jeszcze inni twierdzą, że znaczy to tyle, iż jaźń jest w taki czy inny sposób abstrakcją, czymś, czego istnienie bynajmniej nie podlega wątpliwości przez swą niewidzialność. Można by powiedzieć, że przecież środek narracyjnej ciężkości jest tak samo niewidoczny – i tak samo rzeczywisty. Czy jednak w wystarczającym stopniu?

Kwestię, czy rzeczywiście istnieje jaźń, można przedstawić tak, żeby ją banalnie łatwo rozwiązać, i to zarówno pozytywnie, jak i negatywnie: Czy *my* istniejemy? Oczywiście! Pytanie zakłada swoją własną odpowiedź. (Czym w końcu jest to *ja*, które bez skutku poszukuje jaźni, według Humeʼa?) Czy istnieją byty – albo w naszych mózgach, albo *poza i ponad* nimi – sterujące ciałami, myślące nasze myśli, podejmujące nasze decyzje? Oczywiście, że nie! Pomysł taki jest empirycznym idiotyzmem („papieski neuron” Jamesa) albo metafizycznym, czczym gadaniem („duch w maszynie” Ryleʼa). Kiedy proste pytanie ma dwie odpowiedzi, „Oczywiście, że tak!” i „Oczywiście, że nie!”, warto rozważyć pozycję pośrednią (Dennett 1991a), mimo że z pewnością początkowo będzie sprzeczna z intuicją obu stron – wszyscy zgodzą się, że przeczy takiemu czy innemu oczywistemu faktowi.

1. Jak ludzie przędą jaźń

Poza tym spędzali bodaj mnóstwo czasu, jedząc, pijąc i chodząc na imprezy, a Frensic,

którego wygląd zwykle ograniczał jego przyjemności zmysłowe do wkładania rzeczy w siebie, a nie w innych ludzi, był swego rodzaju smakoszem.

Tom Sharpe, 1977

Pisarz Tom Sharpe sugeruje w tym wesołym, ale niepokojącym fragmencie, że gdy przejść do rzeczy, wszystkie przyjemności zmysłowe polegają na zabawie z własnymi granicami lub granicami innych, i ma rację – jeśli nie jest to cała prawda, to jest to jej część.

Ludzie mają jaźnie. A psy? Homary? Jeśli jaźnie czymkolwiek są, to istnieją. *Teraz* istnieją jaźnie. Były czasy, tysiące (albo miliony lub miliardy) lat temu, gdy ich nie było – a przynajmniej nie na tej planecie. Musi zatem istnieć – logicznie rzecz biorąc – prawdziwa historia, którą można opowiedzieć, o tym, *jak powstały* stworzenia z jaźniami. Historia ta będzie musiała opowiadać – logicznie rzecz biorąc – o procesie (lub serii procesów) związanych z czynnościami czy zachowaniami rzeczy, które jeszcze nie *mają* jaźni – lub też nie *są* jeszcze sobą – ale które w końcu wytworzą, jako coś nowego, istoty posiadające jaźń. W rozdziale 7 widzieliśmy, jak pojawienie się racji było również pojawieniem się granic, granic między „mną" a „resztą świata", rozróżnienia, które muszą robić nawet najprostsze ameby, na swój niewidomy, niewiedzący sposób. Ta minimalna skłonność do odróżniania siebie od innych, by chronić siebie, to jaźń biologiczna, lecz nawet tak prosta jaźń nie jest niczym konkretnym, a jedynie abstrakcją, zasadą organizacji. Co więcej, granice biologicznej jaźni są nieszczelne i niezdefiniowane – kolejny przykład tolerancji Matki Natury na „błąd", jeśli tylko nie jest zbyt kosztowny.

W granicach ludzkiego ciała jest wielu, wielu intruzów, od bakterii i wirusów, poprzez mikroskopijne roztocza, żyjące jak osadnicy na klifie w ekologicznej niszy naszych skór, aż po większe pasożyty – na przykład wstrętne tasiemce. Wszyscy ci intruzi są maleńkimi ochraniaczami samych siebie, jednak niektórzy z nich, jak bakterie zasiedlające nasze układy trawienne, bez których byśmy nie przeżyli, są tak samo kluczowymi członkami zespołu w naszej pogoni za przetrwaniem jak antyciała w naszych systemach immunologicznych. (Jeśli teoria biolożki Lynn Margulis (1970) jest słuszna, to mitochondria działające w niemal każdej komórce naszego ciała są potomkami bakterii, z którymi „my" połączyliśmy siły jakieś dwa miliardy lat temu). Inni intruzi są tolerowanymi pasożytami – najwyraźniej niewartymi wysiłku eksmisji – a jeszcze inni są prawdziwymi wewnętrznymi wrogami, zabójczymi, jeśli ich nie usunąć.

Ta fundamentalna biologiczna zasada odróżniania siebie od świata, wnętrze od zewnętrza, odbija się niesamowitym echem w naszej psychice. Psychologowie Paul Rozin i April Fallon (1987) pokazali w serii fascynujących eksperymentów dotyczących natury *obrzydzenia*, że mamy potężny i niedoceniony ślepy opór na pewne czynności, które rozważane racjonalnie nie powinny być dla nas problemem. Czy mógłbyś na przykład połknąć teraz ślinę w buzi? Ta czynność nie przepełnia cię wstrętem. Ale załóżmy, że proszę cię o przyniesienie czystej szklanki, naplucie do niej i połknięcie śliny ze szklanki. Obrzydliwe! Ale dlaczego? Wydaje się to mieć coś wspólnego z naszym przekonaniem, że gdy coś już wydostanie się z naszego ciała, przestaje być jego częścią – staje się obce i podejrzane – wyrzekło się swojego obywatelstwa i stało się czymś, co należy odrzucić.

Przekraczanie granicy to zatem raczej moment obawy albo, jak pokazał Sharpe, coś, z czego należy szczególnie czerpać radość. Wiele gatunków rozwinęło niesamowite konstrukcje, poszerzające ich granice terytorialne po to, żeby utrudnić ich przekraczanie w zły sposób lub ułatwić dobre. Na przykład bobry budują tamy, a pająki zaplatają pajęczyny. Gdy pająk zaplata pajęczynę, nie musi rozumieć, co robi; Matka Natura po prostu zaopatrzyła jego malutki mózg w potrzebne procedury, by mógł wykonać tę biologicznie niezbędną czynność inżynieryjną. Eksperymenty z bobrami pokazują, że nawet ich niezwykle efektywne praktyki inżynieryjne są

przynajmniej w dużym stopniu wynikiem wewnętrznych popędów i skłonności, których nie muszą rozumieć, aby z nich korzystać. Bobry się uczą i mogą nawet uczyć inne bobry, ale przede wszystkim kierują nimi potężne wewnętrzne mechanizmy kontrolujące to, co behawiorysta B.F. Skinner nazwał „negatywnym wzmocnieniem". Bóbr gwałtownie szuka czegoś – czegokolwiek – aby zatrzymać dźwięk płynącej wody, a w jednym eksperymencie bóbr odnalazł ulgę, zaklejając błotem głośnik, z którego dochodził dźwięk nagranego bulgotania (Wilsson 1974)!

Bóbr chroni swoją zewnętrzną granicę gałązkami i błotem, a jedną ze swoich wewnętrznych granic futrem. Ślimak zbiera wapń z pożywienia i używa go do stworzenia twardej skorupki; krab pustelnik zdobywa gotową skorupę z wapniem, przejmując odrzuconą skorupę innego zwierzęcia, wdzięcznie unikając procesów przyswajania i wydzielania. Różnica nie jest fundamentalna według Richarda Dawkinsa (1982), który zauważa, że rezultatem we wszystkich przypadkach, nazywanych przez niego *fenotypem rozszerzonym*, jest część fundamentalnie biologicznego wyposażenia jednostek, podporządkowanych siłom doboru napędzającym ewolucję.

Definicja fenotypu rozszerzonego nie tylko wychodzi poza „naturalne" granice jednostek, by skorzystać z zewnętrznego wyposażenia, na przykład skorup (oraz wewnętrznego, na przykład zamieszkujących ciało bakterii); często obejmuje inne jednostki tego samego gatunku. Bobry muszą pracować w zespole, aby wybudować jedną tamę, nie mogą zrobić tego same. Termity muszą się skrzykiwać w milionowe grupy, by budować zamki.

Przyjrzyjmy się niesłychanym konstrukcjom architektonicznym australijskich altanników (Borgia 1986). Samce budują złożone altanki, zalotne świątynie z potężnymi głównymi nawami, bogato udekorowane kolorowymi przedmiotami – szczególnie ciemnoniebieskimi, a wśród nich można znaleźć nakrętki od butelek, kawałki kolorowego szkła i inne przedmioty wytworzone przez ludzi – które są zbierane daleko od domu i uważnie układane w altance, aby wywrzeć lepsze wrażenie na samicach, na których im zależy. Altannik, jak pająk, wcale nie musi rozumieć tego, co robi; zwyczajnie ciężko pracuje, nie wiedząc dlaczego, tworząc budowlę, która jest kluczowa dla jego sukcesu jako altannika.

Jednak najdziwniejsze i najwspanialsze konstrukcje w całym świecie zwierzęcym to niesamowite, zawiłe budowle tworzone przez gatunek naczelnych, *homo sapiens*. Każdy normalny przedstawiciel tego gatunku tworzy *jaźń*. Z jego mózgu wysnuwa się sieć słów i czynów i jak inne stworzenia nie musi wiedzieć, co robi; po prostu to robi. Ta sieć chroni go tak jak skorupa ślimaka i zapewnia utrzymanie, jak pajęczyna pająkowi, oraz zwiększa jego szanse na seks, jak altanki altannikowi. W przeciwieństwie do pająka człowiek nie tylko *wysnuwa z siebie* swoją sieć; raczej jak bóbr pracuje ciężko, by zebrać materiały, z których buduje swoją ochronną fortecę. Jak altannik przywłaszcza sobie znalezione przedmioty, które akurat podobają się jemu lub jego partnerowi czy partnerce – w tym wiele takich, które zostały zaprojektowane przez innych do innych celów.

Ta „sieć dyskursów", jak nazwała ją Robyn pod koniec poprzedniego rozdziału, jest tak samo wytworem biologicznym jak każda inna konstrukcja, którą możemy odnaleźć w świecie zwierząt. Pozbawiony tej sieci człowiek jest tak niepełny, jak ptak bez piór czy jak żółw bez skorupy. (Ubrania również są częścią rozszerzonego fenotypu *homo sapiens*, którą można znaleźć niemal w każdej niszy przez nich zamieszkanej. Ilustrowana encyklopedia zoologii nie powinna przedstawiać *homo sapiens* nagiego, tak samo jak nie przedstawia niedźwiedzia czarnego w stroju klauna, jadącego na rowerze).

Organizacja kolonii termitów jest tak wspaniała, że niektórym obserwatorom wydawało się, iż może mieć duszę (Marais 1937). Rozumiemy teraz, że ta organizacja jest zwyczajnie

rezultatem milionów na wpół niezależnych małych podmiotów działających, z których każdy jest automatem, robiąc to, co robi. Organizacja ludzkiej jaźni jest tak wspaniała, że wielu obserwatorom wydawało się, iż każda ludzka istota również ma duszę: życzliwego dyktatora rządzącego z centrali.

W każdym pszczelim ulu czy kolonii termitów z pewnością jest królowa, ale te jednostki są raczej przedmiotami niż podmiotami, podobnie jak klejnoty królewskie, które należy chronić, a nie jak wódz sił zbrojnych – ich królewskie imię tak naprawdę bardziej do nich pasuje dziś niż kiedyś, gdyż o wiele bardziej przypominają królową Elżbietę II niż Elżbietę I. Nie ma pszczoły Margaret Thatcher, nie ma termita George'a Busha, nie ma Gabinetu Owalnego w mrowisku.

Czy nasze ja, nasze maksymalne *jaźnie*, przejawiają tę samą przepuszczalność i giętkość granic jak prostsze ja innych stworzeń? Czy rozszerzamy nasze wewnętrzne granice – granice naszych *jaźni* – aby zawrzeć w nich nasze „rzeczy"? Być może, ogólnie rzecz biorąc, nie, ale z pewnością są momenty, gdy wydaje się to prawdą z psychologicznego punktu widzenia. Na przykład niektórzy ludzie jedynie posiadają samochody i nimi jeżdżą, inni są *kierowcami*; zatwardziały kierowca woli *być* czterokołową istotą zużywającą benzynę, niż być dwunożną spożywającą żywność, a jego użycie zaimka osobowego w pierwszej osobie zdradza tę identyfikację:

Nie skręcam zbyt dobrze w deszczowe dni, bo moje opony się wycierają.

Czasem rozszerzamy więc nasze granice; innymi razy, w odpowiedzi na postrzegane przez nas wyzwania, zarówno prawdziwe, jak i wyimaginowane, pozwalamy, aby nasze granice się kurczyły:

Ja tego nie zrobiłam! To nie ja mówiłam. Tak, słowa wychodziły z moich ust, ale nie rozpoznaję ich jako swoje.

Przypominam o tych typowych wypowiedziach, aby wyrysować podobieństwa pomiędzy naszymi jaźniami a jaźniami mrówek czy krabów pustelników, lecz wypowiedzi te pokazują również najważniejszą różnicę: mrówki i kraby pustelniki nie mówią. Krab pustelnik jest skonstruowany tak, że zajmuje się zdobywaniem skorupy. Moglibyśmy powiedzieć, że jego organizacja *zakłada* skorupę, a stąd, w bardzo słabym sensie, niejawnie *reprezentuje* kraba jako mającego skorupę, ale krab w żadnym silnym sensie nie *reprezentuje siebie* jako mającego skorupę. W ogóle nie zajmuje się reprezentacją siebie samego. Komu miałby się tak reprezentować i dlaczego? Nie musi sobie przypominać o tym aspekcie swojej natury, gdyż jego wewnętrzna konstrukcja rozwiązuje ten problem i nie oczekują tego żadne inne byty. A jak już zauważyliśmy, mrówki i termity wykonują wspólne projekty bez szczegółowo komunikowanych strategii czy zarządzeń.

My natomiast jesteśmy niemal bez przerwy zajęci prezentacją siebie innym i nam samym, a zatem *reprezentowaniem siebie* – w języku i gestach, zewnętrznych i wewnętrznych. Najbardziej oczywista różnica w naszym środowisku wyjaśniająca tę różnicę w naszym zachowaniu to samo zachowanie. Nasze ludzkie środowisko zawiera nie tylko jedzenie i schronienie, wrogów, z którymi walczymy i od których uciekamy, oraz osobniki tego samego gatunku, z którymi możemy połączyć się w pary, ale też słowa, słowa, słowa. Są one potężnymi elementami naszego środowiska, które szybko inkorporujemy, wchłaniamy i wydzielamy, snując je, jak pająk wysnuwa pajęczynę, jako samoobronne ciągi *narracji*. Jak widzieliśmy w rozdziale 7, gdy wpuścimy słowa, te wehikuły memów, rzeczywiście zwykle przejmują one władzę, tworząc *nas* z surowców znajdujących się w naszych mózgach.

Nasza fundamentalna taktyka samoobrony, samokontroli i autodefinicji nie polega na snuciu pajęczyn ani budowie tam, ale opowiadaniu historii, a dokładnie obmyślaniu

i kontrolowaniu historii opowiadanej innym – oraz sobie – o tym, kim jesteśmy. I tak jak pająki nie muszą myśleć, świadomie i celowo, o tym, jak wysnuć pajęczynę, i tak jak bobry, w przeciwieństwie do profesjonalnych ludzkich inżynierów, świadomie i celowo nie planują budowanych przez siebie struktur, tak i my (w przeciwieństwie do *profesjonalnych* ludzkich gawędziarzy) świadomie i celowo nie obmyślamy, jakie narracje opowiadać i w jaki sposób to robić. Nasze bajki są snute, lecz zwykle ich nie snujemy; one snują nas. Nasza ludzka świadomość i nasza narracyjna osobowość/jaźń są jej wytworem, a nie ich źródłem.

Te ciągi czy strumienie narracyjne pochodzą *jakby* z jednego źródła – nie tylko w oczywistym, fizycznym sensie wypływania z jednych ust czy spod pióra lub ołówka, ale subtelniej: ich wpływ na każdą publiczność polega na zachęcaniu jej, aby (spróbowała) przyjęła istnienie jednolitego podmiotu, czyjego słowami są, o którym są: krótko mówiąc, aby przyjęła *środek narracyjnej ciężkości*. Fizycy doceniają ogromne uproszczenia, jakie przynosi założenie środka ciężkości obiektu, pojedynczego punktu, względem którego można obliczyć siły grawitacyjne. My, heterofenomenolodzy, doceniamy ogromne uproszczenie, jakie przynosi założenie środka narracyjnej ciężkości dla ludzkiego ciała snującego narrację. Tak jak jaźń biologiczna, tak ta jaźń psychologiczna czy narracyjna jest kolejną abstrakcją, a nie czymś w mózgu, ale nadal niesłychanie stabilnym i niemal namacalnym atraktorem właściwości, „właścicielem zapisów" wszelkich elementów i cech, które do nikogo nie należą. Kto jest właścicielem twojego samochodu? Ty. Kto jest właścicielem twoich ubrań? Ty. Kto jest zatem właścicielem twojego ciała? Ty! Gdy mówisz:

To jest *moje* ciało.

z pewnością nikt nie twierdzi, że mówisz:

To ciało jest swoim właścicielem.

Co w takim razie masz na myśli? Jeśli twierdzisz, że ani nie jest to przedziwna i bezcelowa tautologia (to ciało jest swoim własnym właścicielem czy coś w tym stylu), ani że jesteś niematerialną duszą czy lalkarzem-duchem, który ma ciało i manipuluje nim w taki sam sposób, w jaki ty masz samochód i nim się posługujesz, cóż innego możesz mieć na myśli?

2. Ile jaźni na jednego klienta?

Myślę, że łatwiej zrozumielibyśmy, co oznacza:

To jest moje ciało.

gdybyśmy potrafili odpowiedzieć na pytanie: W przeciwieństwie do czego? A może w przeciwieństwie do tego?

Nieprawda; jest *moje* i nie lubię go z nikim dzielić!

Gdybyśmy mogli zobaczyć, jak to jest dla dwóch (lub więcej) jaźni rywalizować o kontrolę nad jednym ciałem, moglibyśmy lepiej zrozumieć, czym jest jedna jaźń. Jako naukowcy zajmujący się jaźnią chcielibyśmy przeprowadzić kontrolowane eksperymenty, w których różnicując początkujące warunki, moglibyśmy przekonać się, co takiego musi się stać, w jakiej kolejności oraz przy użyciu jakich środków, aby pojawiła się taka mówiąca jaźń. Czy istnieją warunki, w których toczy się życie, ale nie pojawia się jaźń? Czy istnieją warunki, w których pojawia się więcej niż jedna jaźń? Nie możemy przeprowadzić takich eksperymentów w sposób etyczny, lecz, jak już tyle razy wcześniej, możemy wykorzystać dane wygenerowane przez okrutne eksperymenty prowadzone przez naturę, ostrożnie wyciągając z nich wnioski.

Takim eksperymentem jest osobowość wieloraka, objawiająca się tym, że jedno ludzkie ciało *wydaje się* zamieszkane przez kilka jaźni, z których każda ma zwykle swoje własne imię i autobiografię. Idea osobowości wielorakiej wielu osobom zdaje się zbyt cudaczna

i metafizycznie dziwaczna, aby mogła być prawdziwa – zjawisko „paranormalne", które można odrzucić razem z bliskimi spotkaniami trzeciego stopnia i czarownicami na miotłach. Przypuszczam, że niektórzy z nich popełnili zwykły arytmetyczny błąd: nie zauważyli, że dwie, trzy lub siedemnaście jaźni w jednym ciele nie jest czymś bardziej ekstrawaganckim niż jedna jaźń w jednym ciele. Już z jedną mamy problem!

– Właśnie widziałem samochód z pięcioma jaźniami w środku.

– Co?? Zwariowałeś? Co to za metafizyczny nonsens!

– No cóż, w samochodzie było też pięć ciał.

– No to dlaczego tak nie powiedziałeś? W takim razie wszystko gra.

– A może tylko cztery ciała, albo trzy – ale z pewnością pięć jaźni.

– Co??!!

Normalny układ to jedna jaźń na ciało, ale jeśli może ono mieć jedną jaźń, dlaczego nie więcej niż jedną w nienormalnych warunkach?

Nie próbuję sugerować, jakoby w osobowości wielorakiej nie było nic zadziwiającego czy niesamowicie zagadkowego. Jest to tak naprawdę zjawisko szokujące, ale uważam, że nie dlatego, iż jest wyzwaniem dla naszego postrzegania rzeczy *metafizycznie* możliwych, lecz raczej dlatego, że jest wyzwaniem z jednej strony dla naszego postrzegania tego, co jest możliwe *dla człowieka*, ludzkich granic okrucieństwa i deprawacji, a z drugiej strony dla granic ludzkiej kreatywności. Istnieją obecnie liczne świadectwa na rzecz hipotezy, że są nie garstki czy setki, ale tysiące przypadków osobowości wielorakiej diagnozowanych dziś, a zaburzenie to niemal zawsze powstaje w wyniku długotrwałego znęcania się nad pacjentem w dzieciństwie, zwykle molestowania seksualnego, w odrażająco okrutnym stopniu. Nicholas Humphrey i ja badaliśmy osobowość wieloraką kilka lat temu (Humphrey i Dennett 1989) i stwierdziliśmy, że jest to skomplikowane zjawisko wychodzące daleko poza indywidualne mózgi pacjentów.

Dzieci te często były wychowywane w tak niesamowicie przerażających i dezorientujących warunkach, iż bardziej zadziwia mnie, że w ogóle były w stanie psychicznie przeżyć, niż to, że udało im się siebie ochronić poprzez desperackie zmiany granic. Gdy konfrontują się z przytłaczającym konfliktem i bólem, robią, co następuje: „wychodzą". Tworzą granicę, aby zgroza nie przydarzała się *im*; albo nie przydarza się nikomu, albo innej jaźni, która ma lepsze możliwości utrzymania swojej organizacji w momencie ataku – a przynajmniej pacjenci *mówią*, że tak było, bo tak podpowiada im pamięć.

Jak to możliwe? Jak moglibyśmy zrelacjonować, na poziomie biologicznym, taki proces rozszczepienia? Czy musiała istnieć pojedyncza, pełna osobowość/jaźń, która w jakiś sposób rozdzieliła się niczym ameba? Jak to możliwe, jeśli jaźń nie jest częścią właściwą organizmu czy mózgu, ale – jak już twierdziłem – abstrakcją? Taka odpowiedź na traumę wydaje się ponadto tak twórcza, że z początku skłania do założenia, iż musiała być wytworem swego rodzaju nadzorcy: nadzorczego programu w mózgu czy głównego kontrolera. Przypomnijmy sobie jednak kolonię termitów, która również z początku wydawała się wymagać centralnego nadzorcy, by móc wykonać tak przemyślne projekty.

Przyzwyczailiśmy się do opowieści ewolucyjnych zaczynających się od stanu, w którym pewne zjawisko jeszcze nie istnieje, a prowadzących do stanu, w którym zjawisko jest wyraźnie obecne. Wynalezienie rolnictwa, ubrań, domów i narzędzi, wynalezienie języka, wynalezienie samej świadomości, wcześniejsze powstanie życia na Ziemi. Opowiedzieć można wszystkie te historie. A każda z nich musi przejść coś, co moglibyśmy nazwać otchłanią absolutyzmu. Otchłań ta jest zilustrowana przez poniższy, ciekawy fragment (zapożyczony od Sanforda 1975):

Każdy ssak ma matkę ssaka,

ale istnieje tylko ograniczona liczba ssaków, więc

musiał kiedyś być pierwszy ssak,
co jest zaprzeczeniem pierwszej zasady, więc wbrew pozorom
nie ma czegoś takiego jak ssaki!

Z czegoś trzeba zrezygnować. Ale z czego? Absolutystyczny czy esencjalistyczny filozof ma pociąg do wyraźnych granic, progów, „esencji" i „kryteriów". Dla absolutysty rzeczywiście musiał istnieć pierwszy ssak, pierwsza żyjąca istota, pierwszy moment świadomości, pierwszy podmiot moralny; był to ten wynik skoku, ten radykalnie nowy byt potencjalnie spełniający kryteria, który pierwszy spełnił esencjalne warunki – ukazane przez analizę.

Upodobanie do wyraźnych granic międzygatunkowych było największą intelektualną przeszkodą, z którą musiał mierzyć się Darwin, gdy rozwijał teorię ewolucji (Richards 1987). Przeciwny temu sposobowi myślenia jest rodzaj antyesencjonalizmu, który jest wygodny w niejasnych przypadkach i w braku ściśle wytyczonych granic. Skoro jaźń, umysł, a nawet sama świadomość są wytworami biologicznymi (nie pierwiastkami do znalezienia w układzie okresowym), powinniśmy się spodziewać, że przejścia między nimi oraz innymi zjawiskami będą stopniowe, dyskusyjne, osobliwie poszarpane. Nie oznacza to, że wszystko cały czas podlega zmianie, zawsze stopniowej; oglądane z bliska zmiany, wydające się stopniowymi, zwykle z odleglejszego punktu widzenia wyglądają jak nagłe przejścia między płaskowyżami równowagi (Eldredge i Gould 1972; jednak zob. też Dawkins 1982/2003, s. 136–143).

Istotność tego faktu dla teorii filozoficznych (oraz dla upodobań filozofów) nie jest wystarczająco powszechnie wiadoma. Zawsze były – i zawsze będą – pewne przejściowe elementy, „brakujące ogniwa", quasi-ssaki i inne rzeczy opierające się definicjom, ale faktem jest, że *prawie wszystkie* prawdziwe (w przeciwieństwie do jedynie możliwych) rzeczy w naturze zwykle układają się w zbitki podobnych elementów, a odseparowane są przez ogromne oceany nicości w przestrzeni logicznej. Nie potrzebujemy „esencji" ani „kryteriów", aby znaczenia naszych słów się nie porozjeżdżały; nasze słowa pozostaną na miejscu, dość mocno przywiązane jakby przez grawitację do najbliższej zbitki podobnych do siebie elementów, nawet jeśli istniał – a musiał istnieć – krótki przesmyk, który kiedyś łączył je przez serię stopniowych kroków do jakiejś sąsiedniej zbitki. Ta idea ma zupełnie bezsporne zastosowania w wielu przypadkach. Jednak wielu ludzi, którym jest wygodnie z tym pragmatycznym podejściem do nocy i dnia, życia i świata nieożywionego, ssaka i prassaka, czuje obawę, gdy zaprasza się ich do zastosowania tego samego podejścia do posiadania i nieposiadania jaźni. Uważają, że tutaj bardziej niż gdziekolwiek indziej w naturze musi chodzić o wszystko albo nic oraz o jeden egzemplarz na klienta.

Rozwijana przez nas teoria świadomości podważa te założenia, a osobowość wieloraka to dobra ilustracja tego, jakim wyzwaniem jest dla nich teoria. Przekonanie, że nie mogą istnieć quasi-jaźnie, a ponadto *musi* istnieć pełna liczba jaźni związanych z jednym ciałem – a najlepiej, aby była to liczba jeden – nie jest oczywiste. Innymi słowy, nie jest ono już oczywiste, gdyż rozwinęliśmy dosyć dokładnie konkurencję wobec teatru kartezjańskiego z jego świadkiem czy centralnym nadawaczem sensów. Z jednej strony podważa je osobowość wieloraka, ale z drugiej strony możemy sobie wyobrazić coś innego: co najmniej dwa ciała mające jedną jaźń! Taki przypadek może istnieć w Yorku w Anglii: bliźniaki Greta i Freda Chaplin („Time", 6.04.1981). Te czterdziestoletnie, jednojajowe siostry mieszkające razem w hostelu wydają się działać *jak jedna*; współpracują na przykład, wspólnie podejmując pojedyncze akty mowy i swobodnie kończąc swoje własne zdania czy mówiąc jednocześnie, jedna ułamek sekundy po drugiej. Od lat są nierozłączne, na tyle, na ile mogą to robić bliźnięta niesyjamskie. Niektórzy znający te siostry twierdzą, że naturalną i efektywną taktyką, która sama się nasunęła, jest uważanie *ich* za *nią*.

Nasze postrzeganie pozwala na teoretyczną możliwość nie tylko osobowości wielorakiej,

ale również ułamkowego zaburzenia osobowości (*Fractional Personality Disorder*). Mogłaby istnieć? Dlaczego nie? Zupełnie nie próbuję sugerować, że te bliźniaczki są połączone telepatycznie, przez postrzeganie pozazmysłowe czy inne tajemne więzy. Uważam, że istnieje mnóstwo subtelnych, codziennych sposobów komunikacji i koordynacji (technik rzeczywiście często dobrze opanowanych przez bliźnięta). Skoro te bliźniaczki widzą, słyszą, czują i myślą o właściwie tych samych wydarzeniach podczas całego swojego życia, a niewątpliwie zaczęły, mając mózgi podobnie skłonne do reakcji na takie bodźce, być może nie potrzeba ogromnych kanałów komunikacyjnych, aby były w stanie naprowadzać się na jakiś rodzaj luźnej harmonii. (Poza tym, w jakim stopniu jednolita jest najbardziej wśród nas opanowana osoba?) Powinniśmy wahać się, wyznaczając granice tego rodzaju wypracowanej koordynacji.

W każdym razie, czy nie istniałyby również dwie jasno zdefiniowane jaźnie, każda w jednej bliźniaczce, odpowiedzialne za podtrzymywanie tej dziwnej szarady? Być może, jednak co, jeśli każda z tych kobiet stałaby się tak ofiarna (*selfless*) w swoim poświęceniu dla wspólnej sprawy, że mniej więcej zagubiłaby się w tym projekcie? Jak stwierdził kiedyś poeta Paul Valéry, wspaniale trawestując maksymę swojego rodaka: „Czasem jestem, czasem myślę".

W rozdziale 11 widzieliśmy, że świadomość wydaje się ciągła, choć w rzeczywistości ma luki. Jaźń również mogłaby być tak poprzerywana, popadając w nicość z łatwością, z jaką zdmuchuje się świeczkę, tylko po to, by rozpalić się trochę później, w bardziej pomyślnych warunkach. Czy jesteś tą samą osobą, której przedszkolne przygody tak pobieżnie pamiętasz (czasem wyraźnie, czasem mgliście)? Czy przygody tego dziecka, którego trajektoria przez przestrzeń i czas najwyraźniej była równoczesna z trajektorią twojego ciała, są twoimi przygodami? To dziecko, noszące twoje imię, dziecko, którego nagryzmolony podpis na obrazku przypomina ci o tym, jak *ty* kiedyś pisałaś lub pisałeś swoje imię – jest (było) tym dzieckiem w tobie? Filozof Derek Parfit (1984) porównał osobę do klubu, trochę innego rodzaju ludzkiej konstrukcji, która może przestać istnieć jednego roku, lecz zostać ponownie założona przez niektórych z jej (byłych?) członków parę lat później. Czy byłby to ten sam klub? Być może, jeśli na przykład miałby spisany statut, który by wyraźnie zezwalał na takie luki w egzystencji. Być może jednak nie da się tego stwierdzić. Możemy znać wszystkie fakty, które wyraźnie miałyby wpływ na sytuację, oraz możemy stwierdzić, że były nierozstrzygające o *tożsamości* (nowego?) klubu. W przypadku spojrzenia na jaźnie – czy osoby – które się tu pojawia, jest to poprawna analogia; jaźnie nie są niezależnie istniejącymi duszami-perłami, lecz wytworami procesów społecznych i podobnie do innych takich wytworów mogą podlegać nagłym zmianom statusu. Jedyny „pęd" narastający w trajektorii jaźni czy klubu to stabilność nadana im przez sieć przekonań, które je powołują do życia, a gdy te przekonania zanikają, zanika i pęd, na zawsze lub na jakiś czas.

Warto, byśmy o tym pamiętali, zajmując się kolejnym ulubionym przykładem filozofów, szeroko dyskutowanym zjawiskiem rozszczepienia mózgu. Tak zwane rozszczepienie mózgu jest wynikiem *komisurotomii*, operacji przecinania spoiwa wielkiego mózgu, szerokiego pasma włókien bezpośrednio łączących lewą i prawą półkulę kory mózgowej. Pozostawia to półkule nadal pośrednio połączone, przez różnorodne struktury śródmózgowe, ale jest oczywiście zabiegiem drastycznym, który należy przeprowadzać tylko w sytuacji bez wyjścia. Dostarcza ulgi w niektórych ostrych przypadkach epilepsji, których nie da się leczyć w żaden inny sposób, zapobiegając wewnętrznie generowanym wyładowaniom elektrycznym powodującym napady, które z początkowego ogniska w jednej półkuli przemieszczają się do drugiej. Typowa legenda filozoficzna powiada, że pacjenci z rozszczepieniem mózgu w wyniku operacji mogą zostać „rozszczepieni na dwie jaźnie", ale poza tym nie czują żadnych innych przykrych jej skutków. Najbardziej interesująca wersja tego uproszczenia jest taka, że dwie „strony" oryginalnej osoby –

sztywna, analityczna półkula lewa i rozluźniona, intuicyjna, holistyczna półkula prawa – po operacji są wolne, aby rozbłysnąć jeszcze większą indywidualnością, skoro normalna, bliska współpraca musi zostać zastąpiona mniej ścisłym odprężeniem. Jest to pomysł pociągający, ale widać w nim zdecydowane przerysowanie empirycznych wniosków, które go zainspirowały. W rzeczywistości tylko w niewielkim ułamku przypadków można zaobserwować *jakiekolwiek* teoretycznie frapujące symptomy jaźni wielorakiej. (Zob. np. Kinsbourne 1974; Kinsbourne i Smith 1974; Levy i Trevarthen 1976; Gazzaniga i LeDoux 1978; Gazzaniga 1985; Oakley 1985; Dennett 1985b).

Nie jest zaskoczeniem, że pacjenci z rozszczepieniem mózgu, tak jak ci ze ślepowidzeniem czy z osobowością wieloraką, nie dorastają do oczekiwań filozoficznych, i nie jest to niczyja wina. Nie chodzi o to, że filozofowie (oraz inni interpretatorzy, łącznie z najważniejszymi badaczami) celowo przesadzają w swoich opisach tych przypadków. Starając się zwięźle opisać takie zjawiska, stwierdzają, że ograniczone zasoby codziennego języka nieubłaganie wciągają ich w nazbyt uproszczony model właściciela i ciała, ducha w maszynie, publiczności w teatrze kartezjańskim. Nicholas Humphrey i ja, porównując nasze dokładne notatki dotyczące tego, co wydarzyło się na różnych spotkaniach związanych z pacjentami cierpiącymi na osobowość wieloraką, stwierdziliśmy, że często popadaliśmy, na przekór samym sobie, w zbyt naturalny, ale rzeczywiście zwodniczy język, by opisać, czego właśnie byliśmy świadkami. Thomas Nagel (1971), pierwszy filozof opisujący przypadki pacjentów z rozszczepieniem mózgu, zaprezentował rozsądną i dokładną relację tego zjawiska w sposób, w jaki wówczas je rozumiano, a zauważając trudność przedstawienia spójnego ujęcia, stwierdził: „Być może jest dla nas niepodobieństwem porzucenie pewnych sposobów pojmowania i przedstawiania samych siebie, choćby nie znajdowały najmniejszego wsparcia ze strony badań naukowych" (Nagel 1971/1997, s. 184).

Rzeczywiście jest to trudne, ale nie niemożliwe. Pesymizm Nagela sam w sobie jest przesadzony. Czyż nie udało nam się właśnie uwolnić od tradycyjnych sposobów myślenia? Niektórzy ludzie mogą nie *chcieć* porzucić wersji tradycyjnej. Mogą mieć nawet dobre powody – powody moralne – by próbować zachować mit jaźni jako perły mózgu, poszczególnego konkretu, policzalnej rzeczy, a nie abstrakcji, oraz by odmówić zgody na możliwość quasi-jaźni, pół-jaźni. Lecz jest to zdecydowanie poprawne rozumienie zjawiska rozszczepienia mózgu. Przez krótkie okresy podczas dokładnie zaprojektowanych procedur eksperymentalnych niektórzy z tych pacjentów rozszczepiają się w odpowiedzi na kłopotliwe położenie, tymczasowo tworząc drugi środek narracyjnej ciężkości. Kilka efektów tego rozszczepienia może na czas nieokreślony pozostać we wzajemnie niedostępnych śladach pamięci, ale poza tymi tak naprawdę dosyć prostymi śladami rozszczepienia życie drugiej elementarnej jaźni trwa najwyżej kilka minut, co nie jest czasem wystarczającym na stworzenie takiego rodzaju autobiografii, z jakiej tworzy się dojrzałe jaźnie. (Dotyczy to oczywiście większości dziesiątek fragmentarycznych jaźni rozwiniętych przez pacjentów z osobowością wieloraką; większość z nich zwyczajnie nie ma wystarczająco dużo czasu w ciągu dnia na to, aby stworzyć więcej niż kilka minut rozłącznej biografii na tydzień).

Różnice w osobnych narracjach są życiodajne dla jaźni. Jak zauważył filozof Ronald de Sousa:

Gdy doktor Jekyll przemienia się w pana Hyde'a, jest to rzecz dziwna i tajemnicza. Czy są oni dwiema osobami działającymi na zmianę w tym samym ciele? Ale oto coś jeszcze dziwniejszego: dr Juggle i dr Boggle również działają na zmianę w jednym ciele. *Jednak są do siebie tak podobni, jak bliźnięta jednojajowe!* Wzdrygasz się: dlaczego w takim razie mówią, że zamienili się w kogoś innego? Cóż, czemu nie: jeśli doktor Jekyll może zmienić się w człowieka

tak różnego od siebie, jak pan Hyde, z pewnością musi być o wiele *łatwiej* Juggle'owi i Boggle'owi, który jest dokładnie taki sam.

Potrzebujemy konfliktu lub dużej różnicy, aby odrzucić nasze naturalne założenie, że na jedno ciało przypada najwyżej jeden podmiot działający. [de Sousa 1976, s. 219]

A zatem *jak to jest* być jaźnią prawej półkuli u pacjenta z rozszczepieniem mózgu? Jest to najnaturalniejsze pytanie na świecie[126] i przywołuje niepojęty – i mrożący krew w żyłach – obraz: oto ty w więzieniu prawej półkuli mózgu ciała, którego lewą stronę znasz na wylot (i nadal kontrolujesz), a którego prawa strona jest teraz tak daleka, jak ciało przechodnia. Chcesz powiedzieć światu, jak to jest być tobą, ale nie możesz! Odcięto cię od wszelkiej werbalnej komunikacji, bo nie masz pośrednich linii telefonicznych do radiostacji w lewej półkuli. Starasz się, jak możesz, by zasygnalizować swoje istnienie światu zewnętrznemu, zmuszając swą połówkę twarzy do jednostronnego marszczenia czoła i uśmiechu, a czasem (jeśli jesteś wirtuozem jaźni prawopółkulowej) bazgrząc jedno czy dwa słowa lewą ręką.

To ćwiczenie wyobraźni można by ciągnąć w oczywisty sposób, jednak wiemy, że jest fantazją – taką samą jak urocze opowieści o Piotrusiu Króliku i ich antropomorficznych zwierzęcych przyjaciołach autorstwa Beatrix Potter. Nie dlatego, że „świadomość jest tylko w lewej półkuli", i nie dlatego, że *to niemożliwe*, by ktoś znalazł się w takiej sytuacji, ale po prostu dlatego, że komisurotomia nie pozostawia organizacji, które byłyby zarówno wystarczająco odmienne, jak i silne, aby podtrzymać takie oddzielne jaźnie.

Nie byłoby żadnym wyzwaniem dla mojej teorii jaźni powiedzenie, że jest „logicznie możliwe" istnienie takiej jaźni prawopółkulowej u pacjentów z rozszczepieniem mózgu, gdyż moja teoria mówi, że takie nie jest i wyjaśnia dlaczego: warunki do zebrania takiego narracyjnego bogactwa (i niezależności), które stanowią „dojrzałą" jaźń, nie występują. Moja teoria jest również odporna na twierdzenie – któremu absolutnie nie zamierzam przeczyć – że *mogą istnieć* mówiące króliczki, pająki zapisujące w swoich pajęczynach wiadomości po angielsku czy na przykład melancholijne ciuchcie. Przypuszczam, że mogłyby być, ale ich nie ma – zatem moja teoria nie musi ich wyjaśniać.

3. Nieznośna lekkość bytu

Cokolwiek się wydarza, nieważne gdzie i kiedy, jesteśmy skłonni zastanawiać się, kto lub co jest za to odpowiedzialne. Prowadzi nas to do odkrycia wyjaśnień, których w innym razie nie bylibyśmy sobie w stanie wyobrazić, i pomaga nam to przewidzieć oraz kontrolować nie tylko to, co wydarza się w świecie, ale również to, co dzieje się w naszych umysłach. A co jeśli te same tendencje poprowadzą nas do wyobrażenia sobie rzeczy i przyczyn nieistniejących? Wówczas wymyślimy fałszywych bogów i przesądy, a ich działanie dostrzeżemy w każdym zbiegu okoliczności. Być może to dziwne słowo „ja" – użyte na przykład w zdaniu „Ja właśnie miałem dobry pomysł" – jest odbiciem tej samej tendencji. Jeśli musisz odnaleźć jakąś przyczynę powodującą wszystko, co robisz – no cóż, to coś potrzebuje nazwy. Ty nazywasz to „mną", ja nazywam to „tobą".

Marvin Minsky, 1985, s. 232

Według mojej teorii jaźń nie jest matematycznym punktem, lecz abstrakcją zdefiniowaną przez niezliczoną ilość atrybutów i interpretacji (łącznie z autoatrybucjami i autointerpretacjami) stwarzających biografię żywego ciała, dla którego są środkiem narracyjnej ciężkości. Jako taka jaźń odgrywa szczególnie ważną rolę w toczącej się, poznawczej gospodarce żywego ciała, ponieważ w środowisku, z którego aktywne ciało musi wytworzyć modele umysłowe, nic nie jest

bardziej istotne niż model postrzegania samego siebie przez podmiot działający. (Zob. np. Johnson-Laird 1988; Perlis 1991).

Na samym początku każdy podmiot działający musi wiedzieć, którą rzeczą w świecie jest! Może się to na pierwszy rzut oka wydawać banalne lub niemożliwe. „Jestem sobą!" niekoniecznie brzmi pouczająco, bo cóż więcej jednostka mogłaby chcieć wiedzieć – lub mogłaby odkryć, jeśli jeszcze tego nie wie? Dla prostszych organizmów jest prawdą, że wiedza o sobie nie jest niczym innym od elementarnej mądrości biologicznej przechowywanej pod postacią takich maksym jak: „Gdy jesteś głodny, nie jedz sam!" oraz „Jeśli czujesz ból, to jest on twój!". W każdym organizmie, również ludzkim, potwierdzenie tych podstawowych zasad projektu jest po prostu „wbudowane" – jest częścią konstrukcji układu nerwowego, jak mrugnięcie, gdy coś zbliża się do oka, lub dreszcze, gdy jest zimno. Homar mógłby równie dobrze zjeść szczypce innego homara, ale perspektywa zjedzenia swoich własnych szczypców jest, z korzyścią dla niego, nie do pomyślenia. Jego możliwości są ograniczone i gdy „myśli" on o ruszeniu szczypcami, jego „myśliciel" jest bezpośrednio i odpowiednio połączony z tymi właśnie szczypcami, o których ruszeniu myśli. Jednak w przypadku ludzi (oraz szympansów i być może kilku innych gatunków) istnieje więcej możliwości, a stąd więcej źródeł nieporozumień.

Jakiś czas temu władze nowojorskiego portu eksperymentowały ze wspólnym systemem radarowym dla właścicieli małych łódek. Jedna potężna antena radaru znajdująca się na lądzie tworzyła radarowy obraz portu, który następnie mógł zostać przesłany w postaci sygnału telewizyjnego do właścicieli łódek, ci zaś mogli zaoszczędzić na kosztach radaru, instalując mały telewizor w swojej łódce. Czemu miałoby to służyć? Gdybyśmy zgubili się we mgle i spojrzeli na ekran telewizora, wiedzielibyśmy, że jeden z tych wielu poruszających się na ekranie punktów to my – ale który? Oto przypadek, w którym pytanie: „Jaką rzeczą w świecie jestem?" nie jest ani banalne, ani bez odpowiedzi. Tajemniczość znika dzięki łatwej sztuczce: zróbmy szybko łódką małe kółko; wówczas nasz punkt to ten, który rysuje małe „O" na ekranie – chyba że kilka łódek we mgle próbuje wykonać ten sam test w tym samym czasie.

Metoda nie jest niezawodna, ale zwykle się sprawdza i ilustruje o wiele bardziej ogólną kwestię: aby kontrolować takie wyszukane czynności, jakie wykonują ludzkie ciała, system kontroli ciała (znajdujący się w mózgu) musi być w stanie rozpoznawać szerokie spektrum różnego rodzaju wejść jako informujących go o sobie samym, a jeśli pojawia się dylemat lub wkrada sceptycyzm, jedynym solidnym (choć nie niezawodnym) sposobem na rozstrzygnięcie i odpowiednie przydzielenie informacji jest przeprowadzenie małych eksperymentów: zrób coś i popatrz, co się rusza[127]. Szympans łatwo może się nauczyć sięgać po banany przez dziurę w klatce, jeśli prowadzi ruchy swojego ramienia, obserwując to ramię na ekranie telewizora umieszczonego dość daleko od niego (Menzel i in. 1985). Jest to zdecydowanie niebanalna część rozpoznania samego siebie, zależąca od zauważenia związku między widzianymi ruchami ramienia na ekranie i niewidocznymi, *ale zamierzonymi* ruchami ręką. Co stałoby się, gdyby eksperymentatorzy nadali obrazowi krótkie opóźnienie? Ile twoim zdaniem czasu *tobie* zabrałoby odkrycie, że patrzysz na swoje własne ramię (bez werbalnych podpowiedzi od eksperymentatorów), jeśli obraz byłby opóźniony o, powiedzmy, 20 sekund?

Potrzeba wiedzy o samym sobie wykracza poza problemy z identyfikowaniem zewnętrznych oznak ruchów naszego własnego ciała. Musimy znać nasze wewnętrzne stany, tendencje, decyzje, mocne i słabe strony, a podstawowa metoda uzyskiwania tej wiedzy jest właściwie taka sama: zrób coś, co „widać", i zobacz, co się „rusza". Zaawansowany podmiot działający musi wytworzyć zwyczaje śledzenia zarówno swoich stanów cielesnych, jak i „umysłowych". Jak widzieliśmy, u ludzi zwyczaje te polegają głównie na nieprzerwanym

opowiadaniu i weryfikowaniu historii, z której część jest oparta na faktach, a część jest fikcją. Dzieci praktykują to na głos (przypomnijmy sobie psa Snoopy'ego siadającego na dachu swojej budy i mówiącego do siebie: „Oto as myśliwski drugiej wojny światowej..."). My, dorośli, robimy to bardziej elegancko: w ciszy, milcząco, bez wysiłku śledzimy różnice pomiędzy naszymi fantazjami a „poważnymi" próbami i refleksjami. Filozof Kendall Walton (1973, 1978) oraz psycholog Nicholas Humphrey (1986) z różnych perspektyw pokazali, jak ważny jest teatr, opowiadanie historii oraz bardziej fundamentalne zjawisko udawania w dostarczaniu praktyki ludziom, którzy są świeżo upieczonymi tkaczami jaźni.

W taki sposób tworzymy historię określającą nas samych, uporządkowaną wokół swego rodzaju podstawowego punktu autoreprezentacji (Dennett 1981a). Ten punkt to oczywiście nie jaźń; jest to *reprezentacja* jaźni (a punkt na ekranie radaru odpowiadający wyspie Ellis nie jest wyspą – jest reprezentacją wyspy). To nie wygląd czyni pewien punkt punktem-*ja*, a inny punkt jedynie punktem-*on*, -*ona* czy -*ono*, lecz to, do czego służy. Zbiera i organizuje informacje na *mój* temat w taki sam sposób, w jaki inne struktury w moim mózgu śledzą informacje o Bostonie, Reaganie czy lodach.

A gdzie jest rzecz, *o którą* w twojej autoreprezentacji chodzi? Jest tam, gdzie ty (Dennett 1978b). A *co* jest tą rzeczą? Jest to ni mniej, ni więcej niż twój środek narracyjnej ciężkości.

Powraca Otto:

Problem ze środkami narracyjnej ciężkości jest taki, że nie są one rzeczywiste; są fikcjami teoretycznymi.

Nie jest to problem ze środkami ciężkości; problemem jest ich świetność. Są *znakomitymi* fikcjami, takimi, które chciałby stworzyć każdy. A fikcyjne postaci w literaturze są jeszcze wspanialsze. Weźmy Izmaela z *Moby Dicka*. „Imię moje: Izmael", zaczyna się tekst, a my się na to godzimy. Nie nazywamy Izmaelem tekstu ani Melville'a. Kogo lub co nazywamy Izmaelem? Izmaelem nazywamy Izmaela, wspaniałą, fikcyjną postać, którą można znaleźć na stronach *Moby Dicka*. „Imię moje: Dan", słyszysz z moich ust i się na to godzisz; nie na nazywanie moich ust Danem, lub mojego ciała Danem, ale nazywanie Danem *mnie*, fikcję teoretyczną stworzoną przez... cóż, nie przeze mnie, lecz przez mój mózg, od lat działający w porozumieniu z moimi rodzicami, rodzeństwem i przyjaciółmi.

Może dla ciebie to tak wygląda, ale *ja* jestem zupełnie prawdziwy. Być może ukształtował mnie społeczny proces, do którego właśnie nawiązałeś (musiało tak być, jeśli nie istniałem przed swoim urodzeniem), ale to, co wytwarza ten proces, to *prawdziwa* jaźń, a nie tylko jakaś fikcyjna postać!

Chyba wiem, do czego zmierzasz. Jeśli jaźń nie jest niczym rzeczywistym, to co dzieje się z odpowiedzialnością moralną? Jedna z najważniejszych ról jaźni w naszym tradycyjnym schemacie pojęciowym to miejsce, gdzie spada odpowiedzialność, jak głosił napis na biurku Harry'ego Trumana. Jeśli jaźnie nie są rzeczywiste – nie są *rzeczywiście* rzeczywiste – to czy odpowiedzialność nie będzie się przesuwała i przesuwała, w kółko, na zawsze? Jeśli w mózgu nie ma Gabinetu Owalnego zamieszkanego przez najwyższe władze, u których można odwoływać się od każdej decyzji, wydaje się, że grozi nam kafkowska biurokracja homunkulusów, które zawsze odpowiadają, gdy kwestionuje się ich działanie: „Nie możesz winić mnie, ja tu tylko pracuję". Zadanie skonstruowania jaźni mogącej *brać na siebie* odpowiedzialność jest wielkim społecznym i edukacyjnym projektem. Masz rację, jeśli niepokoi cię zagrożenie jego trwałości. Jednak perła mózgu, „wewnętrznie odpowiedzialne" coś, to żałosne świecidełko wykorzystywane jak amulet w momencie tego zagrożenia. Jedyna nadzieja, jeszcze nie całkowicie płonna, leży w zrozumieniu, naturalistycznie, jak mózg tworzy autoreprezentacje, tym samym wyposażając kontrolowane przez siebie ciało w odpowiedzialną jaźń, gdy wszystko idzie

zgodnie z planem. Warto chcieć mieć wolną wolę i odpowiedzialność moralną, a jak próbowałem pokazać w *Elbow Room: The Varieties of Free Will Worth Wanting* (1984), najlepsza ich obrona polega na porzuceniu beznadziejnie pełnego sprzeczności mitu odrębnej, osobnej duszy.

Ale czy ja nie istnieję?

Oczywiście, że istniejesz. Przecież siedzisz na krześle, czytasz moją książkę i zadajesz pytania. A co ciekawe, twoje obecne ucieleśnienie, choć konieczny warunek twojego zaistnienia, niekoniecznie jest wymagane, aby twoje istnienie zostało przedłużone na czas nieokreślony. Jeśli jesteś duszą, perłą substancji niematerialnej, moglibyśmy „wyjaśnić" twoją potencjalną nieśmiertelność jedynie przez żądanie, by była niedającą się wyjaśnić własnością, niemożliwą do wyeliminowania *virtus dormitiva* duszy. A gdybyś była lub był perłą substancji materialnej, pewną spektakularną grupą atomów w twoim mózgu, twoja śmiertelność zależałaby od fizycznych sił, które trzymają je razem (moglibyśmy zapytać fizyka, czym jest „okres połowiczny trwania" jaźni). Jeśli jednak myślisz o sobie jako o środku narracyjnego ciążenia, twoje istnienie zależy od nieustępliwości narracji (trochę jak w baśniach z tysiąca i jednej nocy, ale z pojedynczą opowieścią), która *teoretycznie* mogłaby przetrwać nieskończenie wiele zmian *nośnika*, być teleportowana z taką łatwością (zasadniczo) jak wieczorne wiadomości i na zawsze przechowywana w postaci czystej informacji. Jeśli to, czym jesteś, to owa organizacja informacji, która nadała strukturę twojemu systemowi kontroli nad ciałem (lub, mówiąc nieco bardziej wyzywająco, jeśli to, czym jesteś, to program uruchomiony na komputerze twojego mózgu), wówczas zasadniczo możesz przetrwać śmierć swojego ciała bez uszczerbku, niczym program może przeżyć zniszczenie komputera, na którym powstała pierwsza jego wersja. Niektórzy myśliciele (np. Penrose 1989/2000) uważają to za potworne i wyjątkowo sprzeczne z intuicją następstwo poglądu, którego tu bronię. Lecz jeśli to, czego najbardziej pragniesz, to potencjalna nieśmiertelność, wówczas mój pogląd jest po prostu bezkonkurencyjny.

Rozdział 14

Wyobrażona świadomość

1. Wyobrażając sobie świadomego robota

Zjawisko ludzkiej świadomości zostało wyjaśnione w poprzednich rozdziałach w kategoriach działania „maszyny wirtualnej", swego rodzaju wytworzonego przez ewolucję (i nadal ewoluującego) programu komputerowego, kształtującego czynności w mózgu. Nie ma teatru kartezjańskiego; są tylko wielokrotne szkice tworzone przez procesy ustalania treści, odgrywające różnorodne, na wpół niezależne role w większej ekonomii mózgu kontrolującej podróż ludzkiego ciała przez życie. Niesamowicie trwałe przekonanie o tym, że istnieje teatr kartezjański, jest wynikiem wielu złudzeń poznawczych, które zostały tu obnażone i wyjaśnione. „*Qualia*" zastąpiono złożonymi stanami dyspozycyjnymi mózgu, a jaźń (znana również jako publiczność w teatrze kartezjańskim, centralny nadawacz sensów czy też świadek) okazała się cenną abstrakcją, fikcją teoretyczną, a nie wewnętrznym obserwatorem lub szefem.

Jeśli jaźń jest „tylko" środkiem narracyjnej ciężkości i jeśli wszystkie zjawiska w ludzkiej świadomości można wyjaśnić „jedynie" jako czynności wirtualnej maszyny realizowane w niesamowicie elastycznie regulowanych połączeniach ludzkiego mózgu, wówczas, zasadniczo, odpowiednio „zaprogramowany" robot z komputerowym mózgiem z krzemu byłby świadomy, miałby jaźń. Innymi słowy, istniałaby świadoma jaźń, której ciało byłoby robotem, a której mózg byłby komputerem. To następstwo mojej teorii jest dla wielu osób oczywiste i bez zarzutu. „*Oczywiście*, że jesteśmy maszynami! Jesteśmy tylko bardzo, bardzo skomplikowanymi, rozwiniętymi maszynami zrobionymi z cząstek organicznych, a nie z metalu i krzemu, i jesteśmy świadomi, więc mogą istnieć świadome maszyny – my". Dla tych czytelników to następstwo jest oczywiste. Mam nadzieję, że to, co okazało się dla nich ciekawe, to różnorodność oczywistych konsekwencji napotkanych po drodze, szczególnie tych pokazujących, ile ze zdroworozsądkowego obrazu kartezjańskiego trzeba zastąpić, gdy dowiadujemy się coraz więcej o rzeczywistej maszynerii mózgowej.

Inni jednak uważają, że konsekwencja mojej teorii mówiąca, iż mógłby zasadniczo istnieć świadomy robot, jest tak niewiarygodna, że w ich oczach równa jest sprowadzeniu mojej teorii do absurdu. Znajomy odpowiedział kiedyś na moją teorię w następujący, szczery sposób: „Ale, Dan, ja nie potrafię wyobrazić sobie świadomego robota!". Niektórzy czytelnicy mogą skłaniać się do przyznania mu racji. Powinni się temu oprzeć, gdyż jest on w błędzie. Błąd ten jest prosty, lecz ukazuje fundamentalne niezrozumienie uniemożliwiające postęp w rozumieniu świadomości. „Wiesz, że to nieprawda – odpowiedziałem. – Często wyobrażasz sobie świadome roboty. Nie chodzi o to, że nie potrafisz go sobie wyobrazić; nie potrafisz sobie wyobrazić tego, *jak* robot mógłby być świadomy".

Każdy, kto widział R2D2 i C3PO w *Gwiezdnych wojnach* lub słuchał HALa w *2001: Odysei kosmicznej*, wyobraził sobie świadomego robota (lub świadomy komputer – to, czy system stoi na własnych nogach jak R2D2, czy jest unieruchomiony jak HAL, nie jest tak naprawdę istotne w naszym zadaniu). Dziecinnie proste jest wyobrażenie sobie strumienia świadomości rzeczy „nieożywionej". Dzieci robią to na okrągło. Wewnętrzne życie mają nie tylko pluszowe misie, ale również lokomotywy. Jodły milcząco stoją w lesie, obawiając się

siekiery drwala, lecz jednocześnie marzą o tym, aby stać się choinką w jakimś miłym, przytulnym domu, otoczoną wesołymi dziećmi. Literatura dziecięca (nie mówiąc o telewizji) pełna jest możliwości wyobrażania sobie świadomych egzystencji takich obiektów. Artyści ilustrujący te fantazje zwykle pomagają wyobraźni dzieci, rysując ekspresyjne twarze na tych fałszywych agentach, ale nie jest to niezbędne. Mówienie – jak HAL – będzie równie dobre pod nieobecność wyrazu twarzy, tworząc iluzję, że ktoś tam jest, że jest to coś, co jest HALem, pluszowym misiem lub lokomotywą.

Oto jednak trudność: wszystko to jest iluzją – a przynajmniej tak się wydaje. Są między nimi różnice. Jest oczywiste, że żaden miś nie jest świadomy, ale nie jest już tak jasne, że świadomy nie mógłby być robot. Oczywiste jest to, że trudno sobie wyobrazić, jak mógłby on istnieć. Skoro mojemu znajomemu trudno było sobie wyobrazić, jak robot mógłby być świadomy, był niechętny, aby wyobrazić sobie robota *mającego mieć* świadomość – choć mógł to zrobić bez żadnego wysiłku. Istnieje ogromna różnica między tymi dwoma wyczynami wyobraźni, a ludzie często je ze sobą mylą. Jest rzeczywiście niesamowicie trudno wyobrazić sobie, jak komputer-mózg robota mógłby być podstawą świadomości. *Jak* skomplikowana masa zdarzeń przetwarzania informacji w kupie układów krzemowych *mogłaby* równać się ze świadomym przeżyciem? Jest jednak równie trudno wyobrazić sobie, jak ludzki mózg zbudowany z materii organicznej mógłby być podstawą świadomości. Jak skomplikowana masa elektrochemicznych oddziaływań między miliardami neuronów *mogłaby* równać się świadomemu przeżyciu? A jednak łatwo wyobrażamy sobie istoty ludzkie jako świadome, nawet jeśli nadal nie potrafimy wyobrazić sobie, *jak* może się to dziać.

Jak mózg może być siedliskiem świadomości? To pytanie zwykle traktowane jest przez filozofów jako retoryczne, co ma sugerować, że odpowiedź na nie jest poza ludzkim pojmowaniem. Głównym celem tej książki jest zburzenie tego przeświadczenia. Przekonywałem, że *można* sobie wyobrazić, jak cały ten złożony ogrom czynności mózgu równy jest świadomemu przeżyciu. Mój argument jest prosty: pokazałem, jak to robić. Okazuje się, że można to sobie wyobrazić, pojmując mózg jako rodzaj komputera. Pojęcia informatyczne pomagają wyobraźni, której potrzebujemy, jeśli mamy wejść na nieznany ląd między naszą fenomenologią w postaci, jaką znamy, czyli „introspekcji", a naszymi mózgami, gdy nauka je przed nami odsłania. Pojmując mózgi jako systemy przetwarzające informacje, możemy stopniowo rozproszyć mgłę i obrać drogę przez nieznane, odkrywając, jak to się dzieje, że mózgi wytwarzają wszystkie te zjawiska. Należy unikać wielu zdradliwych pułapek – kuszących ślepych uliczek, takich jak centralny nadawacz sensów, „wypełnianie" czy „*qualia*" – a z pewnością zostało jeszcze trochę zamieszania i oczywistych błędów w szkicu, który przedstawiłem, ale przynajmniej możemy już zobaczyć, jak wyglądałaby taka ścieżka.

Niektórzy filozofowie głoszą jednak, że to przejście przez nieznane jest absolutnie niemożliwe. Thomas Nagel (1974, 1986) twierdzi, że nie można dostać się do subiektywnego poziomu fenomenologii z obiektywnego poziomu fizjologii. Ostatnio również Colin McGinn uznał, że świadomość ma „ukrytą strukturę", niedostępną zarówno w fenomenologii, jak i fizjologii, a mimo iż mogłaby ona zbudować most nad nieznanym, jest prawdopodobnie na zawsze dla nas niedostępna.

Rodzaj ukrytej struktury, jaką sobie wyobrażam, nie leżałby na żadnym poziomie proponowanym przez Nagela: byłby usytuowany gdzieś między nimi. Ten poziom pośredni, ani fenomenologiczny, ani fizyczny, nie byłby (z definicji) oparty na modelu którejś z tych dwóch stron podziału, a więc nie byłoby niemożliwe, aby nie sięgnął na drugą stronę. Jego opis wymagałby radykalnej innowacji pojęciowej (która, jak argumentowałem, jest dla nas prawdopodobnie nieosiągalna). [McGinn 1991, s. 102–103]

Opis „oprogramowania" czy „maszyny wirtualnej" obrany przeze mnie w tej książce jest dokładnie na pośrednim poziomie opisywanym przez McGinna: nie do końca fizjologiczny czy mechaniczny, a jednak z jednej strony odpowiednio łączący się z maszynerią mózgową, a z drugiej nie do końca fenomenologiczny, ale łączący się odpowiednio ze światem treści, światami (hetero)fenomenologii. Udało nam się! *Wyobraziliśmy* sobie, jak mózg mógłby wytworzyć świadome przeżycie. Dlaczego McGinn uważa, że ta „radykalna innowacja pojęciowa" jest dla nas niedostępna? Czy poddaje różne podejścia, zakładające oprogramowanie w mózgu, rygorystycznej i szczegółowej analizie, ujawniającej ich daremność? Nie. W ogóle ich nie bada. Nawet nie próbuje sobie wyobrazić zakładanego przez siebie poziomu pośredniego; zauważa jedynie, iż wydaje mu się oczywiste, że na tym terenie nie znajdzie się nic.

Owa fałszywa „oczywistość" jest ogromną przeszkodą w postępach rozumienia świadomości. Pojmowanie świadomości jako czegoś, co zdarza się w jakiegoś rodzaju teatrze kartezjańskim, oraz założenie, że nie ma nic złego w tym myśleniu, to najnaturalniejsze rzeczy na świecie. Wydają się oczywiste, dopóki nie spojrzy się z bliska na to, czego można się dowiedzieć z aktywności mózgu, i nie rozpocznie się próby szczegółowego wyobrażenia sobie modelu alternatywnego. Wówczas to, co się dzieje, przypomina dowiedzenie się, jak magik wykonuje swoją sztuczkę. Gdy już dokładnie przyjrzymy się temu, co dzieje się za kulisami, odkryjemy, że tak naprawdę nie widzieliśmy tego, co wydawało nam się, że widzimy na scenie. Ogromna luka między fenomenologią i fizjologią odrobinę się kurczy; widzimy, że niektóre z „oczywistych" cech fenomenologii nie są rzeczywiste: nie istnieje wypełnianie wymysłem; nie ma wewnętrznych *qualiów*; nie ma centralnego źródła znaczenia i działania; nie istnieje magiczne miejsce, gdzie następuje pojmowanie. Tak naprawdę nie ma żadnego teatru kartezjańskiego; różnica między przeżyciami ze sceny a procesami zza kulis traci swój urok. Nadal musimy wyjaśnić mnóstwo niesamowitych zjawisk, ale kilka z najbardziej zaskakujących efektów specjalnych nie istnieje w ogóle, a więc nie potrzebuje wyjaśnienia.

Gdy już zrobimy postęp, wyobrażając sobie, *jak* mózg wytwarza zjawisko świadomości, możemy dokonać kilku drobnych poprawek w zadaniu łatwym: wyobrażaniu sobie, że ktoś lub coś ma świadomość. Możemy myśleć dalej, zakładając jakiegoś rodzaju strumień świadomości, ale nie nadajemy mu już wszystkich tradycyjnych właściwości. Skoro strumień ten został uznany za działanie maszyny wirtualnej realizowane w mózgu, nie jest już „oczywiste", że ulegamy iluzji, wyobrażając sobie strumień na przykład w komputerowym mózgu robota.

McGinn zaprasza czytelników, aby razem z nim się poddali: po prostu nie można sobie wyobrazić, jak oprogramowanie mogłoby sprawić, żeby robot był świadomy. Nawet nie próbuj, mówi. Inni filozofowie promują tę postawę, projektując eksperymenty myślowe, które „działają" właśnie dlatego, że odradzają czytelnikowi próbę dokładnego wyobrażenia sobie, jak oprogramowanie mogłoby to osiągnąć. Co ciekawe, dwa najbardziej znane zawierają aluzję do Chin: chiński naród Neda Blocka (1978) oraz chiński pokój Johna Searle'a (1980/1995, 1982, 1984, 1988)[128]. Oba eksperymenty opierają się na tym samym błędzie wyobraźni, a ponieważ szerzej dyskutowany był eksperyment Searle'a, skoncentruję się na nim. Searle prosi nas o wyobrażenie sobie, że jest zamknięty w pokoju, własnoręcznie symulując ogromny program AI, który przypuszczalnie rozumie po chińsku. Stwierdza, że program zdaje test Turinga, udaremniając wszelkie próby odróżnienia go od prawdziwego użytkownika języka chińskiego przez ludzkich rozmówców. Twierdzi, że z tej jedynie behawioralnej nierozróżnialności nie wynika, iż w chińskim pokoju jest jakakolwiek osoba rozumiejąca po chińsku czy posiadająca chińską świadomość. Searle zamknięty w pokoju i w pośpiechu manipulujący ciągami symboli zgodnymi z programem nie zyskuje w ten sposób żadnej znajomości chińskiego, a w pokoju nie ma również nic, co by ten język rozumiało (to jest „po prostu oczywiste", jak powiedziałby Frank

Jackson).

Ten eksperyment myślowy ma udowodnić niemożliwość tego, co Searle nazywa „silną sztuczną inteligencją", czyli tezy, że „odpowiednio zaprogramowany komputer cyfrowy z odpowiednimi danymi przychodzącymi i wychodzącymi miałby w ten sposób umysł dokładnie w takim sensie, w jakim mają go istoty ludzkie" (Searle 1988a). Różne wersje eksperymentu Searle'a spotkały się w ostatniej dekadzie z eksplozją reakcji, a podczas gdy między innymi filozofowie od początku dostrzegali mankamenty tego eksperymentu rozpatrywanego jako argument logiczny[129], nie można zaprzeczyć, że „wniosek" z niego nadal wydaje się wielu ludziom „oczywisty". Dlaczego? Bo nie wyobrażają sobie oni tego przypadku dostatecznie szczegółowo.

Oto nieformalny eksperyment, który pomoże nam dostrzec, czy moja diagnoza jest poprawna. Najpierw wyobraźmy sobie krótki fragment zwycięskiego dialogu pomiędzy chińskim pokojem a sędzią w teście Turinga. (Dla wygody przetłumaczyłem go z chińskiego na angielski).

SĘDZIA: Słyszałeś o Irlandczyku, który znalazł magiczną lampę? Gdy ją potarł, pojawił się dżin i zaproponował, że spełni jego trzy życzenia. „Chciałbym kufel guinnessa!" – odpowiedział Irlandczyk, a kufel pojawił się natychmiast. Irlandczyk od razu zabrał się za sączenie, a w końcu za łapczywe picie, ale poziom guinnessa w szklance zawsze był w magiczny sposób odnawiany. Po chwili dżinowi skończyła się cierpliwość. „A co z drugim życzeniem?" – zapytał. Irlandczyk odpowiedział: „Cóż, jeszcze raz to samo!".

CHIŃSKI POKÓJ: Bardzo śmieszne. Nie, nie słyszałem tego dowcipu – ale, wiesz, uważam, że etniczne dowcipy są w złym guście. Śmiałem się wbrew sobie, choć tak naprawdę myślę, że powinieneś poszukać innych tematów do rozmowy.

S: No dobrze, ale ja opowiedziałem ci ten dowcip, bo chcę, żebyś mi go wyjaśnił.

ChP: Nudy! Dowcipów nie powinno się wyjaśniać.

S: Mimo to jest to moje pytanie testowe. Czy możesz mi wyjaśnić, jak i dlaczego dowcip jest śmieszny?

ChP: Skoro nalegasz. Widzisz, opiera się on na założeniu, że magicznie napełniająca się szklanka będzie napełniała się zawsze, więc Irlandczyk może mieć tyle piwa, ile zdoła wypić. Nie ma zatem powodu, by chciał drugi, ale jest tak głupi (to fragment, który mi się nie podoba) albo tak upojony alkoholem, że nie dociera to do niego, więc bezmyślnie utożsamiając przyjemność ze swoim pierwszym życzeniem, zamawia dolewkę. Te drugoplanowe założenia nie są oczywiście prawdą, a jedynie częścią tradycji opowiadania dowcipów, zgodnie z którą zawieszamy naszą niewiarę w magię itp. A tak przy okazji, moglibyśmy sobie wyobrazić trochę wymuszony dalszy ciąg, kiedy to okazuje się, że Irlandczyk miał jednak „rację" co do swojego drugiego życzenia – być może planuje urządzić wielką imprezę, a jedna szklanka nie wypełni się na tyle szybko, aby zaspokoić pragnienie gości (nie ma sensu odlewać piwa wcześniej – wszyscy wiemy, że zwietrzałe traci smak). Zwykle nie myślimy o takich komplikacjach, co częściowo wyjaśnia, dlaczego śmiejemy się z dowcipów. Czy to wystarczy?

Nie jest to olśniewający poziom rozmowy, lecz załóżmy, że wystarczył, aby sędzia dał się nabrać. Teraz możemy sobie wyobrazić, że wszystkie te przemowy ChP zostały stworzone przez potężny program, którym pilnie steruje Searle. Trudne do wyobrażenia? Oczywiście, ale skoro Searle twierdzi, że program zdaje test Turinga i skoro ten poziom wyrafinowania konwersacyjnego byłby w jego mocy, to jeśli nie próbujemy sobie wyobrazić złożoności programu tworzącego tego rodzaju konwersację, nie postępujemy zgodnie z poleceniami. Oczywiście musimy sobie również wyobrazić, że Searle nie ma żadnego pojęcia, co robi w chińskim pokoju; widzi tylko zera i jedynki, którymi manipuluje zgodnie z programem. Jest też ważne, że Searle prosi nas o wyobrażenie sobie, iż manipuluje niezrozumiałymi chińskimi

znakami, a nie zerami i jedynkami, gdyż mogłoby to nas skłonić do (nieuzasadnionego) założenia, że ten potężny program działałby jakoś jedynie dzięki „łączeniu" przychodzących chińskich znaków z jakimiś wychodzącymi chińskimi znakami. Żaden taki program by oczywiście nie działał – czy wypowiedzi ChP po angielsku „łączą się" z pytaniami sędziego?

Program, który rzeczywiście byłby w stanie tworzyć wypowiedzi ChP w odpowiedzi na pytania S, mógłby wyglądać jak coś takiego (obserwowany z punktu widzenia wirtualnej maszyny, a nie z poziomu Searle'a). Analiza pierwszych słów, „Słyszałeś o...", aktywowała niektóre z demonów programu odpowiedzialne za wykrywanie dowcipów, które to demony zwołały masę strategii radzenia sobie z fikcją, językiem „intencji drugiej" itp., więc gdy doszło do analizowania słów „magiczna lampa", program miał już niski priorytet wobec reakcji związanych z tym, że nie istnieją magiczne lampy. Aktywowane zostały różnorodne ramy narracyjne związane ze standardowymi dowcipami o dżinach (Minsky 1975) czy też opowieściami (Schank i Abelson 1977), tworząc różne oczekiwania na ciąg dalszy, ale zostały one błyskawicznie zakończone przez puentę, która przywołała zwykły scenariusz („proszenia o dolewkę"), a program zauważył, że było to niespodziewane. W tym samym czasie zostały również zaalarmowane demony wrażliwe na negatywne skojarzenia związane z dowcipami etnicznymi, co ostatecznie doprowadziło do drugiego tematu w pierwszej odpowiedzi ChP... I tak dalej, zdecydowanie dokładniej, niż to przedstawiłem.

Faktem jest, że każdy program, który rzeczywiście byłby w stanie poradzić sobie z przedstawioną konwersacją, musiałby być niebywale prężnym, wyrafinowanym i wielopoziomowym systemem po brzegi wypełnionym „wiedzą o świecie", metawiedzą oraz meta-metawiedzą dotyczącą jego własnych odpowiedzi, prawdopodobnych odpowiedzi ze strony rozmówcy, swoich własnych „motywacji" oraz motywacji rozmówcy i wielu innych rzeczy. Searle oczywiście nie przeczy, że programy mogą mieć tego rodzaju strukturę. Zwyczajnie zniechęca nas do zajęcia się tą kwestią. Jeśli jednak mamy się dobrze spisać w roli osoby wyobrażającej sobie ten przypadek, nie tylko mamy prawo, ale jesteśmy zobowiązani do wyobrażenia sobie, że program Searle'a ręcznie przez niego symulowany ma właśnie tę strukturę – a nawet więcej, jeśli tylko potrafimy sobie to wyobrazić. Jednak mam wrażenie, że nie jest już *oczywiste*, iż nie ma prawdziwego rozumienia dowcipu. *Być może* miliardy czynności wszystkich mających skomplikowaną strukturę części tworzą prawdziwe rozumienie w systemie. Jeśli twoja reakcja na tę hipotezę to stwierdzenie, że nie masz pojęcia, czy takie rozumienie następowałoby w tego rodzaju skomplikowanym systemie, wystarczy ona, by pokazać, że eksperyment myślowy Searle'a zależy potajemnie od twojego wyobrażenia sobie tego przypadku, przypadku nieodpowiedniego, w zbyt uproszczony sposób, i wyciągnięcia z niego „oczywistych" wniosków.

Oto w czym tkwi błąd. Wyraźnie widzimy, że gdyby w tak ogromnym systemie było rozumienie, nie byłoby to rozumienie Searle'a (bo jest on tylko trybikiem w maszynie, nieświadomy kontekstu tego, co robi). Widzimy również, że nic w najmniejszym stopniu podobnego do prawdziwego zrozumienia nie ma w żadnym z fragmentów programowania na tyle małego, aby łatwo go było sobie wyobrazić – cokolwiek by to miało być, jest to jedynie bezmyślna procedura przetwarzania ciągów symboli w inne ciągi symboli według pewnego mechanicznego czy syntaktycznego przepisu. Następnie pojawia się przesłanka: Z pewnością *więcej tego samego*, nieważne ile więcej, nigdy nie złożyłoby się na prawdziwe rozumienie. Ale dlaczego ktoś powinien uważać, że to prawda? Pomyśleliby tak dualiści kartezjańscy, ponieważ uważają, że nawet ludzkie mózgi nie są w stanie same osiągnąć rozumienia; według poglądu kartezjańskiego, aby dokonać cudu rozumienia, potrzebna jest niematerialna dusza. Z drugiej strony, jeśli jesteśmy materialistami przekonanymi, że w taki czy inny sposób nasze mózgi są odpowiedzialne za nasze rozumienie bez cudownej asysty, musimy przyznać, iż prawdziwe

rozumienie jest w jakiś sposób osiągane przez proces składający się z interakcji między ogromem podsystemów, z których żaden sam nie rozumie nic. Argument rozpoczynający się od „ten mały fragment aktywności mózgowej nie rozumie chińskiego ani też nie rozumie go ten większy fragment, którego jest częścią...” prowadzi do niechcianej konkluzji, że nawet aktywność całego mózgu nie jest wystarczająca, aby wyjaśnić rozumienie chińskiego. *Trudno jest sobie wyobrazić*, jak „trochę więcej tego samego” mogłoby tworzyć zrozumienie, ale mamy bardzo dobry powód, by przypuszczać, że tak się dzieje, więc w tym przypadku powinniśmy postarać się bardziej, a nie poddawać się.

W jaki sposób możemy postarać się bardziej? Korzystając z pewnych przydatnych pojęć: pojęcia oprogramowania na poziomie pośrednim, które zostało skonstruowane przez informatyków właśnie po to, aby pomóc nam prześledzić w przeciwnym wypadku niewyobrażalną złożoność dużych systemów. Na poziomach pośrednich widzimy wiele bytów niedostrzegalnych na poziomach bardziej mikroskopowych (takich jak „demony”, do których nawiązywałem wcześniej), którym przypisuje się odrobinę quasi-rozumienia. Wówczas nie jest już tak trudne wyobrażenie sobie, jak „więcej tego samego” może składać się na prawdziwe rozumienie. Wszystkie te demony i inne byty są zorganizowane w ogromne systemy, których czynności organizują się same względem swego własnego środka narracyjnej ciężkości. Searle, pracując w chińskim pokoju, nie rozumie chińskiego, ale nie jest w pokoju sam. Jest tam również System ChP i to *tej* jaźni musimy przypisać jakiekolwiek rozumienie dowcipu.

Ta odpowiedź na przykład Searle’a zwana jest przez niego odpowiedzią systemową. Jest to standardowa odpowiedź ludzi zajmujących się sztuczną inteligencją od samego początku istnienia tego eksperymentu myślowego, ponad dekadę temu, lecz rzadko doceniana jest przez ludzi z innych dziedzin. Dlaczego? Prawdopodobnie dlatego, że nie nauczyli się, jak wyobrażać sobie taki system. Po prostu nie potrafią wyobrazić sobie, jak rozumienie mogłoby być właściwością wyłaniającą się z wielu rozproszonych quasi-rozumień w dużym systemie. Z pewnością nie mogą, jeśli nie próbują, ale jak można im pomóc w tym trudnym zadaniu? Czy jest „oszukiwaniem” myślenie o oprogramowaniu jako o składającym się z quasi-rozumiejących homunkulusów, czy jest to odpowiednia podpórka dla wyobraźni, aby zrozumieć astronomiczną złożoność? Searle popełnia błąd *petitio principii*. Zachęca nas do wyobrażenia sobie, że ten ogromny program składa się z jakiejś prostej architektury tablicy przeglądowej, które bezpośrednio łączy jedne ciągi znaków chińskich z innymi, jak gdyby taki program mógł rzeczywiście zastąpić naprawdę jakikolwiek program. Nie musimy wyobrażać sobie tak prostego systemu i zakładać, że *to* jest program symulowany przez Searle’a, bo żaden taki program nie mógłby dać rezultatów mogących zdać test Turinga, jak zostało stwierdzone. (Podobny argument i jego odparcie znajdziesz w Block 1982 i Dennett 1985).

Poziom złożoności ma znaczenie. Gdyby nie miał, istniałby o wiele prostszy argument przeciwko silnej sztucznej inteligencji: „Hej, spójrz na ten kalkulator. Nie rozumie chińskiego, a każdy wyobrażalny komputer jest tylko gigantycznym kalkulatorem, a zatem żaden komputer nie rozumie chińskiego, co właśnie udało nam się dowieść”. Kiedy weźmiemy pod uwagę złożoność, co zrobić musimy, to wtedy rzeczywiście należy wziąć ją pod uwagę – a nie tylko udawać, że to robimy. Jest to trudne, ale zanim nam się to uda, nie możemy ufać żadnym intuicjom co do tego, co jest „oczywiście” nieobecne. Jak w przypadku Marii badającej barwy według Franka Jacksona, tak eksperyment myślowy Searle’a przekazuje silne, jasne przekonanie tylko wtedy, gdy nie trzymamy się instrukcji. Te pompy intuicji są spaprane; nie wzmacniają naszej wyobraźni, lecz wprowadzają ją w błąd.

A co w takim razie z moimi własnymi pompami intuicji? Co na przykład z robotem Shakeyem, systemem CAD Dla Niewidomych 2.0 czy pacjentami ze ślepowidzeniem,

szkolącymi wzrok na bazie informacji zwrotnych? Czy nie są oni równie podejrzani, równie winni wprowadzania czytelnika w błąd? Z pewnością zrobiłem wszystko, co w mojej mocy, aby opowiedzieć te historie tak, by pokierować wyobraźnią w określony sposób i bez zagłębiania się w złożone szczegóły, moim zdaniem zbędne dla celu, który próbowałem osiągnąć. Jest w tym jednak pewna asymetria: moje pompy intuicji zwykle mają na celu pomóc ci wyobrazić sobie nowe możliwości, a nie przekonać cię, że pewne perspektywy są możliwe. Istnieją wyjątki. Moja odmiana mózgu w naczyniu, rozpoczynająca tę książkę, miała pokazać ci niemożliwość pewnych rodzajów oszustw, a niektóre z eksperymentów myślowych w rozdziale 5 miały pokazać, że *nie dałoby się odróżnić* rewizji treści stalinowskich od orwellowskich, chyba że istniałby teatr kartezjański. Te eksperymenty myślowe zwiększają jednak wyrazistość owej „opozycji"; przykłady kobiety w kapeluszu na przyjęciu, długowłosej kobiety w okularach i inne zostały zaprojektowane tak, aby wyostrzyć właśnie to przeczucie, które następnie próbowałem zdyskredytować argumentami.

Ale niech czytelnik nadal ma się na baczności: moje pompy intuicji, jak wszystkie inne, nie są bezpośrednimi demonstracjami, którymi się wydają; są bardziej sztuką niż nauką. (Dalsze ostrzeżenia przed eksperymentami myślowymi filozofów znajdziesz w Wilkes 1988). Jeśli pomagają nam dostrzec nowe możliwości, które następnie możemy potwierdzić bardziej systematycznymi metodami, jest to osiągnięcie; jeśli zwabiają nas na wygodną ścieżkę, to wielka szkoda. Nawet dobrymi narzędziami można się źle posłużyć, a poradzimy sobie lepiej, jeśli będziemy wiedzieć, jak nasze narzędzia działają.

2. Jak to jest być nietoperzem?

Najbardziej powszechnie cytowany i wpływowy eksperyment myślowy dotyczący świadomości to „Jak to jest być nietoperzem" Thomasa Nagela (1974/1997). Odpowiada na pytania zawarte w tytule, twierdząc, że nie jest możliwe, aby sobie to wyobrazić. Myśl ta jest najwyraźniej dla wielu osób przyjemna; można czasem zobaczyć, że jego artykuł jest cytowany, jakby był osobliwością nad osobliwościami, filozoficznym „wynikiem" – otrzymaną demonstracją faktu, który w końcu musi przyjąć każda teoria.

Nagel dobrze wybrał stworzenia. Nietoperze jako ssaki są do nas wystarczająco podobne, aby podtrzymać przekonanie, że *oczywiście* są świadome. (Gdyby napisał: „Jak to jest być pająkiem?", wielu zastanawiałoby się, skąd pewność, że bycie pająkiem w ogóle jakieś jest). Ale dzięki swojemu systemowi echolokacji – nietoperze mogą „widzieć uszami" – różnią się też od nas wystarczająco, abyśmy mogli poczuć sporą przepaść. Gdyby napisał artykuł zatytułowany „Jak to jest być szympansem", albo jeszcze lepiej – „Jak to jest być kotem?", opinia, że jego pesymistyczna konkluzja jest oczywista, nie byłaby tak jednogłośna. Wielu ludzi jest nadzwyczaj pewnych, że *dokładnie* wiedzą, jak to jest być kotem. (Oczywiście się mylą, chyba że do swojej pełnej uczucia i empatii obserwacji dodali przepastne ilości badań fizjologicznych, ale myliliby się z punktu widzenia Nagela).

Tak czy inaczej, większość ludzi wydaje się dość zadowolona, akceptując „wyniki" Nagela dotyczące niedostępności świadomości nietoperza dla człowieka. Jednak niektórzy filozofowie je zakwestionowali i mieli ku temu dobre powody (Hofstadter 1981; Hardin 1988; Leiber 1988; Akins 1990). Najpierw musimy jasno ustalić, o który wynik chodzi. Nie jest to tylko epistemologiczne czy konfirmacyjne (*evidential*) twierdzenie, że nawet jeśli komuś się udało („przez przypadek") wyobrazić sobie, jak to jest być nietoperzem, nigdy nie bylibyśmy w stanie potwierdzić, że się to udało. Chodzi raczej o to, że my, ludzie, nie posiadamy i nigdy nie nabędziemy środków, reprezentacyjnej maszynerii, aby zaprezentować sobie, jak to jest być

nietoperzem.

To rozróżnienie jest ważne. W rozdziale 12 widzieliśmy podobny wyczyn, wyobrażając sobie, jak musiał się czuć lipszczanin, słuchając którejś z kantat Bacha po raz pierwszy. Ten problem epistemologiczny jest trudny, ale łatwo można sobie z nim poradzić dzięki zwykłym metodom badania. Dojście do tego, jakie rodzaje przeżyć mieli lipszczanie i jak różniłyby się one od naszych przeżyć Bacha, to kwestia badań historycznych, kulturowych, psychologicznych i może fizjologicznych. Możemy łatwo zorientować się w niektórych z tych kwestii, łącznie z niektórymi najbardziej uderzającymi różnicami z naszych własnych przeżyć, gdybyśmy jednak mieli wprowadzić się w tę samą sekwencję stanów świadomości, którymi cieszyłaby się taka osoba, spotkalibyśmy się z prawem malejących zysków. Zadanie to wymagałoby od nas ogromnej przemiany – zapomnienia wielu rzeczy, o których wiemy, porzucenia pewnych skojarzeń i nawyków, a nabycia nowych. Możemy użyć naszych badań „trzecioosobowych", aby powiedzieć, czym byłyby te przemiany, ale rzeczywiste poddanie się im wiązałoby się ze strasznymi kosztami izolacji od naszej współczesnej kultury – niesłuchaniem radia, nieczytaniem o wydarzeniach politycznych i społecznych itd. Nie ma potrzeby się do tego wszystkiego uciekać, aby dowiedzieć się czegoś o świadomości lipszczan.

To samo jest prawdą, jeśli chcemy sobie wyobrazić, jak to jest być nietoperzem. Powinniśmy zainteresować się tym, co możemy wiedzieć o świadomości nietoperza (jeśli możemy wiedzieć cokolwiek), a nie tym, czy możemy zmienić nasze umysły czasowo lub trwale w umysły nietoperza. W rozdziale 12 podważyliśmy założenie, że istnieją „wewnętrzne" właściwości – *qualia* – tworzące *to, jak to jest* mieć takie czy inne świadome przeżycia, a jak zauważa Akins (1990), nawet *gdyby* istniały szczątkowe, niedyspozycyjne, nierelacyjne cechy przeżyć nietoperza, bliskie zapoznanie się z nimi, bez poznania dostępnych faktów dotyczących systematycznej struktury systemu percepcyjnego i zachowania nietoperza, nie dałoby nam wiedzy, jak to jest być nietoperzem. Możemy dowiedzieć się bardzo dużo o tym, co znaczy bycie nietoperzem, ale ani Nagel, ani nikt inny nie dał nam dobrego powodu do tego, aby uwierzyć, że jest coś interesującego czy teoretycznie ważnego, co jest dla nas niedostępne.

Nagel twierdzi, że żadna wiedza trzecioosobowa nie mogłaby nam powiedzieć, jak to jest być nietoperzem, a ja kategorycznie temu zaprzeczam. Jak możemy rozstrzygnąć tę dysputę? Angażując się w coś, co zaczyna się jak dziecięca zabawa – zabawa, w której jedna osoba wyobraża sobie, jak to jest być *x,* a druga próbuje zademonstrować, że jest coś nie tak z tym szczególnym ćwiczeniem w heterofenomenologii.

Oto kilka prostych ćwiczeń na rozgrzewkę:

A: Oto pluszowy miś myśli sobie, jak miło byłoby zjeść trochę miodu na śniadanie.

B: Źle. Pluszowy miś nie ma możliwości odróżnienia miodu od innych rzeczy. Nie ma żadnych działających organów zmysłowych ani nawet żołądka. Miś jest wypchany bezwładnym materiałem. Bycie misiem nie jest niczym.

A: Oto Bambi, jelonek podziwiający piękny zachód słońca, a nagle jasne, pomarańczowe słońce przypomina mu o kurtce złego myśliwego!

B: Źle. Jelenie nie rozróżniają barw (choć mogą mieć rodzaj dychromatycznego widzenia). Czegokolwiek świadome są jelenie (jeśli są świadome czegokolwiek), nie rozróżniają barw takich jak pomarańczowy.

A: Oto nietoperz Billy, postrzegający za pomocą echolokacji, że rzecz lecąca w jego kierunku to nie jego kuzyn Bob, ale orzeł z rozpostartymi skrzydłami i szponami przygotowanymi do ataku!

B: Poczekaj – mówiłeś, że jak daleko jest orzeł? Echolokacja nietoperzy działa jedynie na kilka metrów.

A: No cóż... Orzeł był tylko dwa metry od niego.

B: Okej, teraz trudniej powiedzieć. Jakie dokładnie są granice rozdzielczości echolokacji nietoperza? Czy służy ona w ogóle do identyfikowania obiektów, czy jest jedynie alarmistą i detektorem ofiar? Czy nietoperz byłby w stanie odróżnić skrzydła rozpostarte od złożonych jedynie przy użyciu echolokacji? Wątpię, ale będziemy musieli zaprojektować jakieś eksperymenty, aby to potwierdzić, no i oczywiście przygotować eksperymenty, aby odkryć, czy nietoperze są w stanie śledzić i rozpoznawać swoich krewnych. Niektóre ssaki to potrafią, lecz mamy powody twierdzić, że inne są zupełnie tych kwestii nieświadome.

Rodzaj badań zasugerowanych w tym ćwiczeniu przybliżyłby nam strukturę percepcyjnego i behawioralnego świata nietoperza, więc moglibyśmy uporządkować narracje heterofenomenologiczne według stopnia realizmu, odrzucając te, które uznawały lub zakładały talenty do rozróżniania czy dyspozycje do reagowania w sposób widoczny, nieuzasadniony w ekologii i neurofizjologii nietoperza. Dowiedzielibyśmy się na przykład, że nietoperzom nie przeszkadzają emitowane przez nie głośne piski, które wywołują echo, gdyż mają mięsień wyraźnie skonstruowany tak, by zamykał ich uszy dokładnie w momencie emitowania pisku, podobnie do urządzeń pozwalających wrażliwym systemom radarowym unikać zagłuszenia swoimi własnymi sygnałami. Te kwestie były już wielokrotnie badane, więc możemy powiedzieć jeszcze więcej na przykład o tym, dlaczego nietoperze używają innych schematów częstotliwości dla swoich pisków w zależności od tego, czy szukają ofiary, zbliżają się do celu, czy atakują (Akins 1989, 1990).

Gdy napotykamy narracje heterofenomenologiczne, do których odrzucenia żaden krytyk nie potrafi znaleźć żadnych dobrych podstaw, powinniśmy je zaakceptować – ostrożnie, oczekując na kolejne odkrycia – jako trafne ujęcia tego, jak to jest być danym stworzeniem. Tak w końcu traktujemy siebie nawzajem. Proponując traktowanie nietoperzy i innych potencjalnych przedmiotów interpretacji w ten sam sposób, nie *przesuwam* ciężaru dowodu, lecz ekstrapoluję zwykły, ludzki ciężar dowodu na inne byty.

Moglibyśmy wykorzystać te badania, aby położyć kres iluzjom o świadomości nietoperzy. *Wiemy*, że wspaniała książka dla dzieci Randalla Jarrella *The Bat-Poet* (1963) jest fantazją, gdyż wiemy, że nietoperze nie mówią! Tezy na temat fenomenologii nietoperzy, które są w sposób mniej oczywisty fantastyczne, opierają się na mniej oczywistych, ale nadal publicznych faktach o ich fizjologii i zachowaniu. Badania takie powiedziałyby w dużej mierze to, czego mógłby, a czego nie mógłby być świadomy nietoperz w różnych warunkach, pokazując nam możliwości w jego układzie nerwowym, jeśli chodzi o reprezentowanie tego czy tamtego, a eksperymentalnie można byłoby sprawdzić, czy nietoperz rzeczywiście korzysta z tych informacji, aby zmieniać swoje zachowanie. Trudno sobie wyobrazić, ile tak naprawdę można nauczyć się z tego rodzaju badań, dopóki rzeczywiście się nimi nie zainteresujemy. (Niesamowicie szczegółowe, wstępne badanie tego, jak to jest być na przykład koczkodanem, przeprowadzili Cheney i Seyfarth – zob. ich książka *How Monkeys See the World*, 1990).

Rodzi to oczywisty zarzut: badania takie powiedziałyby nam bardzo dużo o organizacji mózgu i przetwarzaniu informacji przez nietoperze, jednak pokazałyby nam jedynie to, czego nietoperze *nie* są świadome, pozostawiając całkowicie otwartą kwestię tego, czego, jeśli w ogóle czegokolwiek, świadome *są*. Jak wiemy, wiele procesów przetwarzania informacji w układach nerwowych jest całkowicie nieświadomych, więc te metody badań bynajmniej nie wykluczą hipotezy, że nietoperze są... latającymi zombi, stworzeniami, którymi bycie nie jest niczym! (Wilkes 1988, s. 224, zastanawia się, czy echolokacja nie jest rodzajem ślepowidzenia).

A, no i wyszedł nietoperz z worka. Dyskusja ta wydaje się podążać w rzeczywiście złowrogim kierunku, więc musimy stanąć jej na drodze. Richard Dawkins (1986/1994)

w pouczających rozważaniach na temat konstrukcji echolokacji u nietoperzy klarownie przedstawia wyłaniający się obraz.

Ze zjawiska Dopplera korzystają radary policyjne „łapiące" kierowców przekraczających dozwoloną prędkość. [...] Porównując częstotliwość wysyłanego sygnału z częstotliwością powracającego echa, policjant – a raczej jego *automatyczne urządzenie radarowe* – może wyliczyć prędkość każdego samochodu. [...] Porównując wysokość swojego pisku z wysokością powracającego echa, nietoperz (a raczej jego *komputer pokładowy w mózgu*) może więc, przynajmniej teoretycznie, wyliczyć, jak szybko porusza się w stronę drzewa. [Dawkins 1986/1994, s. 62, podkr. moje – D.C.D.]

Nasuwa się pytanie: Czy w nietoperzu jest coś w takiej relacji do jego „komputera pokładowego" (działającego bez cienia świadomości), w jakiej jest policjant do „automatycznego urządzenia radarowego"? Policjant nie musi świadomie wyliczać przesunięcia dopplerowskiego, ale ma świadome przeżycia, gdy sczytuje z urządzenia „120 km/h" wypisane jasnymi symbolami na ekranie LED. Wówczas wie, że musi wsiąść na motor i włączyć syrenę. Możemy w sposób wiarygodny przypuszczać, że nietoperz również świadomie nie oblicza przesunięcia dopplerowskiego – zajmuje się tym jego komputer – ale czy nie pozostaje wówczas jakaś nieobsadzona rola wewnątrz nietoperza, coś w rodzaju świadomego policjanta, świadka doceniającego (świadomie) dane wyjściowe z komputera badającego przesunięcie dopplerowskie? Zauważmy, że łatwo moglibyśmy zamienić policjanta na automatyczne urządzenie jakoś zapisujące tablice rejestracyjne przekraczającego prędkość pojazdu, sprawdzające nazwisko i adres właściciela, a następnie wysyłające mu mandat. Nie ma nic szczególnego w zadaniu policjanta, co nie mogłoby zostać dokonane bez świadomości czegokolwiek. Wydaje się, że to samo odnosi się do nietoperza. Nietoperz mógłby być zombi. Byłby zombi – jak sugeruje ten wywód – chyba że znajdowałby się w nim wewnętrzny obserwator reagujący na wewnętrzną prezentację w sposób, w jaki policjant reaguje na migoczące czerwone światełko na swoim urządzeniu.

Nie wpadnij w tę pułapkę. To nasza stara nemezis, publiczność w teatrze kartezjańskim. *Twoja* świadomość faktu polega na tym, że twój mózg zamieszkuje wewnętrzny podmiot działający, przed którym mózg dokonuje prezentacji, więc nieznalezienie go w mózgu nietoperza nie zagrażałoby tezie o jego świadomości ani też naszemu twierdzeniu, że potrafimy powiedzieć, jaka jest jego świadomość. Aby zrozumieć świadomość nietoperza, musimy po prostu zastosować do niego te same zasady, jakie stosujemy do siebie.

Cóż jednak mógłby zrobić nietoperz, co byłoby na tyle niesamowite, aby przekonać nas, że mamy do czynienia z prawdziwą świadomością? Mogłoby się wydawać, że bez względu na to, jak wyrafinowanych użytkowników informacji wyjściowych ustawimy za dopplerowskim przetwornikiem w nietoperzu, nie ma przekonującego, zewnętrznego, „trzecioosobowego" powodu, aby uznać, że nietoperz ma świadome przeżycia. Gdyby potrafił on na przykład mówić, wytworzyłby tekst, z którego moglibyśmy ulepić heterofenomenologiczny świat, i dałoby to nam dokładnie te same podstawy do tego, aby stwierdzić, że ma świadomość, jak w przypadku każdej osoby. Ale, jak już zauważyliśmy, nietoperze nie mówią. Potrafią jednak zachowywać się na różne niewerbalne sposoby mogące zapewnić jasne podstawy do opisu ich świata heterofenomenologicznego, lub też – jak nazwał to pionierski badacz Jakob von Uexküll (1909) – ich *Umwelt und Innenwelt*, czyli „wokół świata" i świata wewnętrznego.

Heterofenomenologia bez tekstu nie jest niemożliwa, a jedynie trudna (Dennett 1988a, 1988b, 1989a, 1989b). Jedna z gałęzi heterofenomenologii zwierzęcej zwana jest „etologią kognitywną" i stanowi próbę zrozumienia zwierzęcych umysłów przez badanie ich zachowania w terenie i eksperymentowanie na nim. Możliwości i trudności tego rodzaju badania dobrze

przedstawiają Cheney i Seyfarth (1990), Whiten i Byrne (1988) oraz Ristau (1991), w tomie poświęconym Donaldowi Griffinowi, pionierskiemu badaczowi echolokacji u nietoperzy i twórcy dziedziny etologii kognitywnej. Jedną z frustrujących trudności napotykanych przez tych badaczy jest to, że wiele z doświadczeń, które chcieliby przeprowadzić, okazuje się zupełnie niepraktycznych bez obecności języka; nie można po prostu *przygotować* badanych (i wiedzieć o tym, że się ich przygotowało) w sposób wymagany przez te doświadczenia bez rozmowy z nimi (Dennett 1988a).

Nie jest to jedynie epistemologiczny problem heterofenomenologów; sama trudność w tworzeniu odpowiednich warunków eksperymentalnych w środowisku naturalnym mówi coś bardzo fundamentalnego o umysłach istot bez języka. Pokazuje, że ekologiczne sytuacje tych zwierząt nigdy nie dały im *szansy* na rozwinięcie (przez ewolucję, uczenie się czy jedno i drugie) wielu zaawansowanych czynności umysłowych kształtujących nasze umysły, więc możemy być właściwie pewni, że te zwierzęta nigdy ich nie rozwinęły. Na przykład zastanówmy się nad koncepcją *sekretu*. Sekret to nie tylko coś, co wiesz, a czego nie wiedzą inni. Aby mieć sekret, musisz wiedzieć, że inni go nie znają, i musisz być w stanie ten fakt kontrolować. (Jeśli jako pierwszy lub pierwsza zobaczysz zbliżający się w panice tłum, być może wiesz coś, o czym nie wiedzą inni, ale nie na długo; nie możesz *zatrzymać* takiej uprzywilejowanej informacji jako sekretu). Ekologia behawioralna gatunku musi być raczej szczególnie ustrukturyzowana, aby sekrety mogły odgrywać jakąkolwiek rolę. Antylopy w swoich stadach nie mają sekretów ani sposobu, aby jakieś uzyskać. Zatem antylopa prawdopodobnie nie ma większych możliwości na uknucie tajemnego planu niż na policzenie do stu lub podziwianie barw zachodzącego słońca. Nietoperze robiące stosunkowo samotne wypady, podczas których być może są w stanie rozpoznać izolację od rywali, spełniają jeden z warunków potrzebnych do posiadania sekretu. Czy mają również interesy, którym mogłoby posłużyć utrzymywanie czegoś w tajemnicy? (Co sekretnie miałby knuć małż? Siedzieć w błocie, chichocząc sam do siebie?) Czy nietoperze mają również zwyczaje zachowywania ostrożności lub oszukiwania podczas polowań, które mogłyby zostać przystosowane do bardziej rozwiniętych czynności utrzymywania tajemnicy? Tak naprawdę jest wiele pytań tego rodzaju, a gdy je zadamy, sugerują dalsze badania i eksperymenty. Struktura umysłu nietoperza jest tak samo dostępna jak struktura jego układu trawiennego; w celu badania którejś z nich należy systematycznie porównywać badania ich treści i badania świata, z którego te treści zostały zaczerpnięte, zwracając uwagę na metody i cele tego czerpania.

Wittgenstein powiedział kiedyś: „Gdyby lew mógł mówić, nie potrafilibyśmy go zrozumieć" (1958/2000, s. 313). Ja uważam przeciwnie, że jeśli lew potrafiłby mówić, miałby on umysł tak różny od zwykłego mózgu lwiego, że mimo możliwości zrozumienia go bez problemu niewiele dowiedzielibyśmy się od niego o zwykłych lwach. Jak widzieliśmy we wcześniejszych rozdziałach, język odgrywa ogromną rolę, nadając strukturę umysłowi człowieka, i nie powinniśmy zakładać, że umysł istoty nieposiadającej języka – ani żadnej potrzeby posiadania go – będzie miał tę samą strukturę. Czy oznacza to, że zwierzęta nieposiadające języka „zupełnie nie są świadome" (jak głosił Kartezjusz)? To pytanie zawsze pojawia się w takim momencie jako rodzaj pełnego niedowierzania wyzwania, nie powinniśmy jednak czuć się zobligowani do odpowiedzenia na nie w takiej formie. Zauważmy, że zakłada ono coś, od czego usilnie próbowaliśmy uciec: przypuszczenie, że świadomość jest niezwykłą właściwością w rodzaju „wszystko albo nic", dzielącą wszechświat na dwie diametralnie różne kategorie: rzeczy, które ją mają (rzeczy, którymi bycie jest jakieś, jak powiedziałby Nagel), oraz te, które jej nie mają. Nawet w naszym własnym przypadku nie możemy wyrysować linii oddzielającej nasze świadome stany umysłowe od nieświadomych. Naszkicowana przez nas teoria świadomości

pozwala na wiele wersji architektury funkcjonalnej, a podczas gdy obecność języka oznacza szczególnie dramatyczny wzrost zakresu wyobraźni, wszechstronności oraz samokontroli (aby wspomnieć jedynie o niektórych bardziej oczywistych mocach joyce'owskiej maszyny wirtualnej), moce te nie mają *dalszej* mocy włączania jakiegoś specjalnego wewnętrznego światła, które w innym przypadku pozostałoby wyłączone.

Gdy wyobrażamy sobie, jak to jest być istotą bez języka, naturalnie zaczynamy od swojego własnego doświadczenia, a większość tego, co przychodzi nam wówczas do głowy, trzeba dostosować (głównie rezygnując z wielu rzeczy). Świadomość w rodzaju tej, jaką cieszą się takie zwierzęta, jest w porównaniu do naszej obcięta. Na przykład nietoperz nie tylko nie może się zastanawiać, czy dziś jest piątek; nie może nawet zastanawiać się nad tym, czy jest nietoperzem; w jego poznawczej strukturze taki namysł nie ma roli do odegrania. Podczas gdy nietoperz, tak jak nawet skromny homar, ma jaźń biologiczną, nie ma jako takiej własnej jaźni godnej tego miana – nie ma środka narracyjnej ciężkości, a co najwyżej taki, który jest bez znaczenia. Nie ma słów na końcu języka ani niczego nie żałuje, nie ma złożonych pragnień, nostalgicznych wspomnień, żadnych wielkich planów, żadnych przemyśleń dotyczących tego, jak to jest być kotem, ani nawet tego, jak to jest być nietoperzem. Ta lista odrzuceń byłaby tanim sceptycyzmem, gdybyśmy nie mieli pozytywnej teorii empirycznej, na której można ją oprzeć. Czy twierdzę, iż udowodniłem, że nietoperze *nie mogą* mieć tych stanów umysłowych? No cóż, nie, ale nie potrafię również udowodnić, że grzyby *nie mogą* być śledzącymi nas, międzygalaktycznymi statkami kosmicznymi.

Czy nie jest to okropnie antropocentryczne uprzedzenie? A poza tym, co z głuchoniemymi? Czy oni nie są świadomi? Oczywiście, że są – ale nie przechodźmy zbyt szybko do ekstrawaganckich konkluzji o ich świadomości, powodowani źle pojmowanym współczuciem. Gdy osoba głuchoniema nabywa język (szczególnie język migowy, najbardziej naturalny język, którego taka osoba może się nauczyć), rodzi się stuprocentowy ludzki umysł, w sposób wyraźnie różniący się od umysłu osoby słyszącej, ale zdolny do złożonego namysłu i mający siłę tworzenia – być może w większym stopniu niż osoba słysząca. Jednak bez naturalnego języka umysł osoby głuchoniemej jest straszliwie zredukowany. (Zob. Sacks 1989/1998, szczególnie bibliografia). Jak napisał filozof Ian Hacking (1990) w recenzji książki Sacksa, „potrzeba żywej wyobraźni, aby mieć choćby najmniejsze pojęcie o tym, czego brakuje niesłyszącemu dziecku". Nie oddajemy przysługi osobom głuchoniemym, wyobrażając sobie, że pod nieobecność języka cieszą się wszystkimi umysłowymi przyjemnościami, które znają osoby słyszące, ani nie oddajemy przysługi zwierzętom innym niż ludzie, próbując zaciemniać dostępne fakty dotyczące ograniczeń w ich mózgach.

A oto, jak wielu z was bardzo chciałoby zaznaczyć, podtekst, który od dłuższego czasu starał się wydostać na powierzchnię: wielu ludzi boi się wyjaśnienia świadomości, bo boją się, że jeśli uda nam się tego dokonać, utracimy nasze podstawy moralne. Uważają, że może i potrafimy wyobrazić sobie świadomy komputer (lub świadomość nietoperza), ale *nie powinniśmy próbować*. Jeśli popadniemy w ten zły nawyk, zaczniemy traktować zwierzęta jak nakręcane zabawki, niemowlęta i głuchoniemych, jak gdyby byli pluszowymi misiami, czy też – jeszcze pogarszając sprawę – roboty, jak gdyby były prawdziwymi ludźmi.

3. Uwaga i znaczenie

Tytuł tego podrozdziału zapożyczyłem z artykułu pt. *Minding and Mattering* Marian Stamp Dawkins, która przeprowadziła dokładne badania moralnych następstw heterofenomenologii zwierzęcej. (Jej wcześniejsza praca została zrelacjonowana przez nią

w książce *Animal Suffering: The Science of Animal Welfare*, 1980). Jak zauważa, nasze postawy moralne wobec innych zwierząt są pełne sprzeczności.

Wystarczy, że pomyślimy o różnych rodzajach zwierząt, aby ujawniły się niespójności. Organizujemy demonstracje przeciwko zabijaniu młodych fok grenlandzkich, ale nie ma podobnych kampanii przeciwko zabijaniu szczurów. Wielu ludzi nie ma problemu z jedzeniem świń czy owiec, ale nie wyobrażają sobie zjedzenia psa czy konia. [Dawkins 1987, s. 150]

Dawkins wskazuje na dwie tendencje w tym chaosie: możliwość rozumowania i możliwość odczuwania cierpienia. Kartezjusz przypisywał dużą wagę temu, że zwierzęta inne niż ludzie nie potrafią rozumować (przynajmniej w sposób, w jaki rozumują ludzie), co wywołało słynną odpowiedź ze strony brytyjskiego filozofa utylitaryzmu, Jeremy'ego Benthama: „Jednakże dorosły koń lub pies jest bez porównania rozumniejszym i zdolniejszym do porozumienia się zwierzęciem od dziecka mającego dzień czy tydzień, czy nawet miesiąc życia. Przypuśćmy jednak, że jest inaczej. Co by to pomogło? Należy pytać nie o to, czy zwierzęta mogą *rozumować* ani czy mogą *mówić*, lecz czy mogą *cierpieć*?" (Bentham 1789/1958, s. 419–420). To wydają się przeciwne punkty odniesienia w kwestiach moralnych, ale jak mówi Dawkins, „nadanie wartości etycznej istotom mogącym odczuwać cierpienie doprowadzi w końcu do tego, że będziemy cenić zwierzęta, które są bystre. Nawet jeśli zaczniemy od odrzucenia kryterium rozumowania Kartezjusza, to myślące racjonalnie zwierzęta najprawdopodobniej mają możliwość odczuwania cierpienia" (s. 153).

Racje po temu kryją się w prezentowanej tu teorii świadomości. Cierpienie nie jest kwestią wchodzenia w jakieś niewyrażalne, ale same w sobie okropne stany, lecz posiadania życiowych nadziei, planów, projektów zniszczonych przez okoliczności uniemożliwiające realizację pragnień, udaremnionych intencji – czymkolwiek one są. Idea cierpienia jako poddającego się wyjaśnianiu przez jakieś wewnętrzne właściwości – nazwijmy je „okropnościami" – jest tak beznadziejna, jak idea wyjaśniania radości jako obecności wewnętrznej wesołości. Zatem przypuszczalna niedostępność, ostateczna niewiedza o czyimś cierpieniu jest tak samo mylna jak inne fantazje o wewnętrznych *qualiach*, które zdemaskowaliśmy, choć jest w sposób bardziej oczywisty szkodliwa. Wynika z tego – a porusza to czułą, intuicyjną strunę – że możliwość odczuwania cierpienia jest możliwością posiadania wyartykułowanych, dalekosiężnych, wysoce specyficznych pragnień, oczekiwań i innych wyrafinowanych stanów umysłowych.

Ludzie nie są jedynymi istotami na tyle mądrymi, by cierpieć; koń i pies Benthama pokazują swoim zachowaniem, że mają wystarczającą złożoność umysłową, aby rozróżnić – i wartościować – całkiem spore spektrum bólów i innych udręk, nawet jeśli jest to wąska paleta w porównaniu z zakresem możliwości cierpienia ludzkiego. Inne ssaki, szczególnie małpy człekokształtne, słonie i delfiny, najwyraźniej mają o wiele szerszy zakres tych odczuć.

W zamian za konieczność znoszenia całego tego cierpienia mądre stworzenia mogą się dobrze bawić. Trzeba mieć ekonomię poznawczą z budżetem na eksplorację i autostymulację, w której jest miejsce na powracające pragnienia, umożliwiające dobrą zabawę. Robisz w tym kierunku pierwszy krok, gdy twoja architektura pozwala ci cieszyć się znaczeniem: „Chwila, przecież ja to uwielbiam!". Płytkie wersje tej mocy konstrukcyjnej są widoczne u pewnych wyższych gatunków, ale potrzeba bujnej wyobraźni i czasu wolnego – coś, na co większość gatunków nie może sobie pozwolić – aby uzyskać szerokie spektrum przyjemności. Im jest ono szersze, tym bogatsze są detale, dokładniej rozróżnialne są pragnienia, a tym gorzej jest, gdy pragnienia zostają udaremnione.

Możesz zapytać, dlaczego istotne jest to, że pragnienia jakiegoś stworzenia zostają zaprzepaszczone, jeśli nie są to pragnienia *świadome*. Odpowiadam: Dlaczego byłoby bardziej

istotne, gdyby były one świadome – zwłaszcza jeśli świadomość byłaby własnością, jak uważają niektórzy, która na zawsze wymyka się badaniu? Dlaczego zaprzepaszczone nadzieje „zombi" mają mniejsze znaczenie niż zaprzepaszczone nadzieje osoby świadomej? To pewna sztuczka z lustrami, którą trzeba ujawnić i odrzucić. Mówisz, że świadomość jest tym, co ważne, a następnie trzymasz się doktryn o świadomości, które w sposób systematyczny uniemożliwiają nam zrozumienie tego, *dlaczego* jest ona ważna. Postulowanie specjalnych, wewnętrznych właściwości, które są nie tylko prywatne i wewnętrznie cenne, ale również niemożliwe do potwierdzenia i zbadania, jest po prostu obskurantyzmem.

Dawkins pokazuje, w jaki sposób poddające się badaniu różnice – jedyne, które mogłyby być istotne – mogą być testowane doświadczalnie, a wystarczy kilka szczegółów, aby pokazać, ile możemy się dowiedzieć nawet ze zwykłych eksperymentów z raczej mało atrakcyjnymi gatunkami.

Kury trzymane na wybiegu lub w dużych zagrodach większość czasu poświęcają na dziobanie podściółki i w związku z tym przypuszczałam, że jej brak w klatkach na fermach masowych może sprawiać, że cierpią. Okazało się, że gdy postawiłam je przed wyborem między klatką z drucianą podłogą oraz z podściółką, którą mogły dziobać, wybrały tę drugą. Wchodziły do malutkiej klatki (na tyle małej, że ledwo co mogły się w niej obrócić), jeśli był to jedyny sposób na dostanie się do podściółki. Nawet ptaki całe życie hodowane w klatkach, które nigdy wcześniej nie doświadczyły podściółki, wybierały klatkę, która miała ją na podłodze. Choć było to sugestywne, nie wystarczyło. Musiałam pokazać nie tylko, że kury preferują ściółkę, ale również, że ich preferencja była na tyle silna, że można powiedzieć, że być może cierpią, gdy jej nie mają.

Kurom zaoferowano wówczas trochę inny wybór. Tym razem musiały wybrać pomiędzy klatką z drucianą podłogą, w której znajdowało się jedzenie i woda, a klatką ze ściółką bez jedzenia i wody. [...] Rezultat był taki, że spędzały dużo czasu w klatce ze ściółką, a o wiele mniej w klatce z drucianą podłogą, mimo że było to jedyne miejsce, gdzie mogły jeść i pić. Wówczas wprowadzono komplikację. Kury musiały „pracować", aby poruszać się między klatkami. Musiały albo przeskoczyć z korytarza, albo przecisnąć się przez zasłonę z czarnego plastiku. Więc zmiana z jednej klatki na drugą była teraz obarczona kosztami. [...] Kury nadal spędzały tę samą ilość czasu w klatce z podłogą drucianą i z jedzeniem, tak jak wówczas, gdy nie miały problemu, by do niej wejść. Ale prawie w ogóle nie spędzały czasu w klatce ze ściółką. Zwyczajnie nie wydawały się gotowe na pracę czy zapłatę za przedostanie się do klatki ze ściółką. [...] Zupełnie inaczej, niż się spodziewałam, ptaki wydawały się mówić, że ściółka tak naprawdę nie była dla nich istotna. [Dawkins 1987, s. 157–159]

Wyciąga wniosek, że „cierpienie przeżywane przez emocjonalny umysł ujawnia się u zwierząt, u których jest on wystarczająco racjonalny, aby móc wpłynąć na warunki sprawiające, że cierpią", a następnie pisze, że „jest również prawdopodobne, że organizmy bez umiejętności zrobienia czegoś, aby odsunąć się od źródła cierpienia, nie rozwinęłyby umiejętności doświadczania cierpienia. Nie byłoby żadnego ewolucyjnego celu, aby drzewo, którego gałęzie są odcinane, mogło cierpieć w milczeniu" (Dawkins 1987, s. 159). Jak widzieliśmy w rozdziale 7 (zob. również rozdział 3 przyp.), należy zachować ostrożność, tworząc takie ewolucyjne argumenty o funkcji, gdyż historia odgrywa ogromną rolę w ewolucji, ale jednocześnie może płatać figle. Jednak pod nieobecność pozytywnych podstaw do przypisywania cierpienia lub pozytywnych podstaw do przypuszczania, że są one z takiego czy innego powodu systematycznie ukrywane, powinniśmy wywnioskować, iż cierpienia nie ma. Nie powinniśmy bać się, że ta surowa zasada doprowadzi nas do lekceważenia innych istot. Nadal zapewnia szerokie pole do pozytywnych konkluzji: wiele zwierząt, ale nie wszystkie, ma możliwość

odczuwania cierpienia w stopniu znaczącym. Bardziej przekonujący argument na rzecz humanitarnego traktowania innych zwierząt opiera się na przyznaniu, że istnieją ogromne różnice w stopniu doświadczania cierpienia, a nie na usilnym upowszechnianiu nieznośnego dogmatu o uniwersalności i równości zwierzęcego bólu.

Być może jest to odpowiedź na obiektywne pytanie o obecność bądź brak cierpienia, ale nie wyjaśnia moralnych odczuć wzburzonych przez perspektywę wyjaśnienia świadomości w tak bezduszny, mechanistyczny sposób. Stawka jest wyższa.

Mam farmę w Maine i uwielbiam to, że w moich lasach żyją niedźwiedzie i kojoty. Bardzo rzadko widuję te zwierzęta czy nawet tylko ślady ich obecności, lecz po prostu lubię wiedzieć, że tam są, i byłbym nieszczęśliwy, gdybym dowiedział się, że odeszły. Nie czułbym również, że taka strata zostałaby zrekompensowana, gdyby moi przyjaciele zajmujący się sztuczną inteligencją zapełnili las mnóstwem bestii-robotów (choć pomysł ten, gdy go sobie szczegółowo wyobrażam, jest urokliwy). Jest dla mnie ważne, że są tam dzikie stworzenia, potomkowie dzikich stworzeń, żyjące tak blisko mnie. Podobnie zachwyca mnie, że w Bostonie i okolicach odbywają się koncerty, których nie tylko nie wysłuchuję, ale *o których* w ogóle nie wiem.

Są to fakty szczególnego rodzaju. Są one dla nas ważne po prostu dlatego, że ważnym dla nas elementem środowiska jest nasze przekonanie o środowisku. A skoro nie jest łatwo nas oszukać, abyśmy wierzyli w pewne sądy, gdy zniknęło dla nich uzasadnienie, ważne jest dla nas, aby te przekonania były *prawdziwe*, nawet jeśli my sami nie będziemy dostrzegali żadnych bezpośrednich na to dowodów. Jak każda inna część środowiska, nasze przekonanie o nim może być kruche, złożone z części połączonych ze sobą zarówno przez historyczne wypadki, jak i dobrze zaprojektowane ogniwa. Zastanów się na przykład nad delikatną częścią naszego przekonania o środowisku dotyczącego tego, co stanie się z naszymi ciałami po śmierci. Niewielu ludzi wierzy, że dusza zamieszkuje ciało po śmierci – nawet osoby wierzące w duszę w *to* nie wierzą. A mimo to niewielu z nas, jeśli w ogóle ktokolwiek, tolerowałoby „reformę" zachęcającą ludzi do pozbywania się martwych ciał ich najbliższych przez wyrzucanie ich w plastikowych workach do śmieci czy też w inny sposób nieceremonialne ich usunięcie. Dlaczego? Nie dlatego, że wierzymy, iż zwłoki mogą doświadczyć jakiegoś upokorzenia. Zwłoki nie mogą być upokorzone bardziej niż kłoda. A mimo to pomysł taki jest szokujący, odrażający. Dlaczego?

Powody są skomplikowane, ale możemy poruszyć teraz kilka prostych kwestii. Osoba to nie tylko ciało; osoba *ma* ciało. Te zwłoki to ciało starego, kochanego Jonesa, środka narracyjnej ciężkości, który zawdzięcza swoją realność zarówno wspólnym wysiłkom wzajemnej interpretacji heterofenomenologicznej, jak i ciału, w którym nie ma teraz życia. Granice Jonesa nie są identyczne z granicami jego ciała, a jego interesy, dzięki ciekawej ludzkiej praktyce snucia jaźni, mogą sięgać poza podstawowe, biologiczne interesy, które zapoczątkowały tę praktykę. Traktujemy jego zwłoki z szacunkiem, gdyż jest to ważne do podtrzymania przekonania o środowisku, w którym wszyscy żyjemy. Jeśli na przykład zaczniemy traktować zwłoki jak śmieci, może to zmienić sposób, w jaki traktujemy prawie-zwłoki – tych, którzy nadal żyją, ale umierają. Jeśli nie popełnimy błędu, przedłużając rytuały i praktyki tego rodzaju daleko za próg śmierci, umierający (i ci, którzy się o nich troszczą) będą musieli stawić czoło lękowi, zniewagom, *możliwościom*, które mogą ich urazić. „Złe" traktowanie zwłok być może nie zrani bezpośrednio żadnej umierającej osoby, a z pewnością nie zrani zwłok, ale gdyby stało się to powszechną praktyką, a przy tym znaną powszechnie (co jest nieuniknione), znacząco zmieniłoby to przekonanie o środowisku otaczającym osobę umierającą. Ludzie zaczęliby wyobrażać sobie wydarzenia następujące po ich odejściu inaczej, niż robią to teraz, i w sposób, który mógłby być szczególnie przygnębiający. Być może bez żadnego dobrego powodu, ale co

z tego? Jeśli ludzie mają być przygnębieni, to samo w sobie jest to dobrym powodem, aby nie przyjmować takiego podejścia.

Istnieją więc pośrednie, lecz nadal godne uznania, usankcjonowane, znaczące powody, by dobrze traktować zwłoki. Nie potrzebujemy żadnej mitologii o czymś szczególnym, co miałoby zamieszkiwać nasze zwłoki i nadawać im przywileje. *Mógłby* to być użyteczny mit, który należałoby rozpowiedzieć wśród mniej obytych, ale myślenie, że my, lepiej poinformowani, musimy podtrzymywać tego rodzaju mity, byłoby skrajnie protekcjonalne. Tak samo istnieją dobre racje po temu, aby traktować wszelkie żywe zwierzęta troskliwie i z dbałością. Racje te są w pewien sposób niezależne od faktów związanych z tym, jakiego rodzaju ból czują zwierzęta. Bardziej bezpośrednio zależą one od faktu, że różne przekonania są obecne w naszej kulturze i są dla nas istotne bez względu na to, czy *powinny* takie być. Skoro są ważne, są ważne. Jednak racjonalność przekonań o środowisku – fakt, że głupawe czy bezpodstawne przekonania zwykle na dłuższą metę wygasają pomimo przesądów – sugeruje, że coś, co jest istotne teraz, może nie pozostać takie na zawsze.

Ale jak przewidzieliśmy w rozdziale 2, teoria radykalnie atakująca powszechne środowisko przekonań ma realny potencjał, by wyrządzać zło, zadawać cierpienie (na przykład u ludzi, którzy szczególnie troszczą się o zwierzęta, bez względu na to, czy to, co przydarza się tym zwierzętom, można nazwać cierpieniem). Czy oznacza to, że powinniśmy zaprzestać badania tych kwestii ze strachu przed otwarciem puszki Pandory? Mogłoby to być uzasadnione, gdybyśmy byli w stanie przekonać samych siebie, że nasze obecne przekonania o środowisku, powodowane mitami czy też nie, są zdecydowanie akceptowalnymi moralnie, dobroczynnymi przekonaniami, lecz przyznaję, że oczywiście tak nie jest. Ci, którzy martwią się ewentualnymi kosztami tego nieproszonego oświecenia, powinni porządnie przyjrzeć się kosztom obecnych mitów. Czy naprawdę uważamy, że to, z czym teraz mamy do czynienia, jest warte ochrony przez jakiś kreatywny obskurantyzm? Czy na przykład uważamy, że ogromne zasoby powinny zostać zarezerwowane, aby zachować urojone perspektywy odnowionego życia umysłowego osób głęboko otępiałych, podczas gdy nie ma zasobów, by wzmocnić desperackie, ale z pewnością nie urojone, oczekiwania biednych? Mity o świętości życia czy o świadomości mają swoje dobre i złe strony. Mogą być użyteczne w tworzeniu barier (przeciwko eutanazji, karze śmierci, aborcji, jedzeniu mięsa), robiących wrażenie na osobach bez wyobraźni, lecz za cenę obraźliwej hipokryzji czy kuriozalnego samooszukiwania się wśród osób bardziej oświeconych.

Bariery absolutne, jak Linia Maginota, rzadko spełniają funkcję, dla której zostały zaprojektowane. Kampania, którą swego czasu prowadzono przeciwko materializmowi, okryła się już wstydem, a kampania przeciwko „silnej sztucznej inteligencji", choć z równie dobrymi intencjami, jako konkurencyjne może zaproponować jedynie najbardziej wyświechtane modele umysłu. Z pewnością lepiej byłoby spróbować docenić *nie*absolutne, *nie*wewnętrzne, *nie*dychotomiczne podstawy zagadnień moralnych, mogące współistnieć z naszą rosnącą wiedzą o wewnętrznym funkcjonowaniu tej najbardziej zadziwiającej maszyny, mózgu. Moralne argumenty *po obu stronach* kwestii na przykład kary śmierci, aborcji, jedzenia mięsa czy eksperymentowania na zwierzętach innych niż ludzie podniosą się na wyższy, bardziej właściwy poziom, kiedy w sposób całkowity odrzucimy mity, które w każdym przypadku są nie do obronienia.

4. Wyjaśnienie czy obalenie świadomości?

Kiedy dowiadujemy się, że jedyną różnicą pomiędzy złotem i srebrem jest liczba cząstek subatomowych w ich atomach, możemy poczuć się oszukani albo się złościć – ci fizycy czegoś

nas pozbawili: złotość zniknęła ze złota; odrzucili samą srebrność srebra, którą tak doceniamy. A kiedy wyjaśniają, jak odbijanie i absorpcja promieniowania elektromagnetycznego odpowiadają za barwy i widzenie barwne, wydaje się, że zaniedbują coś, co jest dla nas najważniejsze. Oczywiście jednak coś musi być „pominięte" – w innym wypadku nie moglibyśmy rozpocząć wyjaśniania. Pomijanie czegoś nie jest cechą złego wyjaśnienia, ale wyjaśnienia udanego.

Jedynie teoria wyjaśniająca zdarzenia świadome przez odwołanie do zdarzeń nieświadomych mogłaby w ogóle wyjaśnić świadomość. Jeśli twój model tego, jak ból powstaje wskutek czynności mózgowych, nadal ma pudełko z napisem „ból", to nawet jeszcze nie zaczynasz wyjaśniać, czym ból jest, a jeśli twój model świadomości rozwija się dobrze do magicznego momentu, w którym musisz powiedzieć „a następnie staje się cud", nie zaczynasz nawet wyjaśniania, czym jest świadomość.

Prowadzi to niektórych ludzi do stwierdzenia, że świadomości nigdy nie będzie można wyjaśnić. Ale dlaczego miałaby ona być jedyną rzeczą, której nie można wyjaśnić? Ciała stałe, ciecze i gazy mogą zostać wyjaśnione w kategoriach rzeczy, które same w sobie nie są ciałami stałymi, cieczami i gazami. Życie również może być wyjaśnione w kategoriach rzeczy, które same w sobie żywe nie są – a wyjaśnienie nie pozostawia rzeczy żyjących bez życia. Przypuszczam, że iluzja tego, iż świadomość jest wyjątkowa, pojawia się, gdyż nie udaje nam się zrozumieć najważniejszej cechy dobrego wyjaśnienia. Błędnie myśląc, że wyjaśnienie czegoś nas pozbawia, chcemy ocalić coś, co w innym przypadku zostałoby zagubione, wkładając to z powrotem do obserwatora jako *quale* – czy jakąś inną „wewnętrznie" cudowną właściwość. Psyche staje się ochronną spódnicą, pod którą mogą ukryć się wszystkie te ukochane kociaki. Mogą istnieć *powody*, by sądzić, że świadomość nie może być wyjaśniona, ale mam nadzieję, iż udało mi się pokazać, że istnieją dobre *racje*, by sądzić, że może.

Moje wyjaśnienie świadomości bynajmniej nie jest pełne. Można by nawet powiedzieć, że jest dopiero początkiem, bo *jest* początkiem, gdyż dzięki niemu pryska czar magicznego kręgu idei, sprawiających, że wyjaśnienie świadomości wydaje się niemożliwe. Nie zastąpiłem teorii metaforycznej, teatru kartezjańskiego, teorią *nie*metaforyczną („dosłowną, naukową"). Tak naprawdę jedynie zastąpiłem jedną rodzinę metafor i obrazów inną, zamieniając teatr, świadka, centralnego nadawacza sensów, wymysł na oprogramowanie, maszyny wirtualne, wielokrotne szkice, pandemonium homunkulusów. Można powiedzieć, że to jedynie wojna na metafory – ale one nie są „tylko" metaforami; są narzędziami myślenia. Nikt nie może myśleć o świadomości bez nich, więc należy wyposażyć się w najlepszy dostępny zestaw tych narzędzi. Zobaczmy, co z ich pomocą zbudowaliśmy. Czyż można by sobie teraz bez nich poradzić?

Załącznik A (dla filozofów)

W książce są miejsca, w których pośpiesznie i bez komentarza przechodzę nad istotnymi, filozoficznymi sporami, czy, innymi słowy, skandalicznie nie wypełniam standardowych obowiązków filozofa akademickiego. Filozofowie, którzy czytali manuskrypt tej książki, pytali o te luki. Ich pytania dotyczą kwestii, które mogą nie interesować osób niebędących filozofami, ale zasługują na odpowiedzi.

Chyba wystrychnąłeś nas na dudka w rozdziale 11, w dialogu z Ottonem, gdy pokrótce przedstawiasz „przeczucia" jako akty mowy bez aktora i mowy, a następnie rewidujesz swoją własną karykaturę, zastępując przeczucia „wydarzeniami ustalania treści" bez żadnego dalszego wyjaśnienia. Czy nie jest to kluczowy ruch w całej tej teorii?

Rzeczywiście. Jest to główny punkt styczności z drugą połową mojej teorii umysłu, teorią treści czy intencjonalności przedstawionej ostatnio w *The Intentional Stance*. W książce jest wiele innych miejsc, gdzie opieram się na tej teorii, ale to punkt, w którym jest ona chyba kluczowa. Bez tej teorii treści byłoby to miejsce, w którym moja własna teoria powiedziałaby „a potem staje się cud". Moja główna strategia zawsze była taka sama: najpierw rozwinąć koncepcję treści, która jest *niezależna od świadomości* i *bardziej fundamentalna od* niej – koncepcję treści, która równo traktuje wszelkie nieświadome ustalanie treści (w mózgach, w komputerach, w „rozpoznawaniu" przez ewolucję właściwości konstrukcji wytworzonych przez dobór) – a następnie zbudować na tej podstawie koncepcję świadomości. Najpierw treść, potem świadomość. Dwie połowy *Brainstorms* podsumowują tę strategię, ale gdy połówki teorii rosły, przerosły jeden tom. Ta książka jest trzecim urzeczywistnieniem tego przedsięwzięcia. Ta strategia jest oczywiście kompletnym przeciwieństwem wizji Nagela i Searle'a, którzy, każdy na swój sposób, obstają przy tym, że świadomość należy traktować jako podstawę. Powodem, dla którego tak prędko zostawiłem za sobą ten prawdziwie istotny temat w rozdziale 11, jest to, że nie wiedziałem, jak można by skrócić setki stron analizy i argumentacji, które poświęciłem teorii treści, do czegoś zarówno precyzyjnego, jak i przystępnego. Jeśli zatem myślisz, że na tych stronach spłatałem ci figla, proszę cię o lekturę bardziej rozbudowanej wersji w innych publikacjach, które przywołuję w bibliografii.

Wydaje się jednak istnieć napięcie – jeśli nie oczywista sprzeczność – między dwiema połówkami twojej teorii. Nastawienie intencjonalne zakłada (bądź wspiera) racjonalność, a stąd jedność podmiotu działającego – systemu intencjonalnego – podczas gdy model wielokrotnych szkiców całkowicie sprzeciwia się tej centralnej jedności. Jak zatem według ciebie należy rozumieć umysł?

Wszystko zależy od tego, z jak daleka patrzysz. Im bliżej podejdziesz, tym lepiej zobaczysz niejednorodność, wielorakość i konkurowanie. Głównym źródłem mitu o teatrze kartezjańskim jest w końcu leniwe ekstrapolowanie nastawienia intencjonalnego *do oporu*. Traktowanie złożonej, poruszającej się jednostki jako podmiotu o jednym umyśle genialnie ułatwia dostrzeżenie wzorca we wszystkich czynnościach; ta taktyka przychodzi nam naturalnie i jest prawdopodobnie preferowana nawet genetycznie jako sposób postrzegania i myślenia. Gdy jednak staramy się tworzyć nauki o umyśle, musimy nauczyć się powstrzymywać i przekierowywać te nawyki myślowe, rozbijając podmiot o jednym umyśle na minipodmioty i mikropodmioty (bez jednego szefa). Wówczas dostrzeżemy, że wiele *pozornych* zjawisk świadomego przeżywania jest błędnie opisanych przez tradycyjną, unitarną taktykę. To napięcie amortyzują naciągane utożsamienia elementów heterofenomenologicznych (rozumianych z tradycyjnej perspektywy) ze zdarzeniami ustalania treści w mózgu (rozumianych z nowej

perspektywy).

Filozofowie często zwracają uwagę na idealizacje taktyki tradycyjnej, ale już nie tak często się na nie godzą. Na przykład sporo literatury filozoficznej skupia się na trudności logiki zwrotnych stanów przekonań i wiedzy, a zapoczątkowane to zostało przez Jaakko Hintikkę. Jedna z istotnych idealizacji formalizacji Hintikki, jak sam to przyznał, polegała na tym, że zdania podlegające zaprezentowanej przez niego logice „muszą powstać *przy jednej i tej samej okazji*. [...] Pojęcie zapominania nie daje się zastosować w granicach takiej okazji" (Hintikka 1962, s. 7). Waga takiego ograniczenia, twierdzi Hintikka, nie zawsze była dostrzegana – a zwykle gubiła się we mgle kontrowersji późniejszych sporów. Uznaje on, że to kwantowanie „okazji" jest koniecznym uproszczeniem, wymaganym do jego formalizacji zdroworozsądkowych pojęć przekonań i wiedzy; umiejscawia treść w jakiejś chwili i w ten sposób ustala tożsamość sądu, o którym jest mowa. Uważam tu, że ta sztuczna indywiduacja „stanów" i „momentów" jest jedną z cech zmieniających te pojęcia psychologii potocznej w fantazje, gdy staramy się je odnieść do złożoności procesów w mózgu.

Czy w takim razie uważasz, że są świadome przeżycia? Czy jesteś zwolennikiem teorii identyczności, materialistą eliminacyjnym, funkcjonalistą, instrumentalistą?

Rzeczywiście nie wyrażam mojej teorii w postaci jednego, formalnego, odpowiednio skwantyfikowanego sądu. Wypełnianie formuły ∀*x x jest świadomym przeżyciem wtedy i tylko wtedy, gdy*... i bronienie jej przeciw zaproponowanym kontrprzykładom nie jest dobrą metodą rozwijania teorii świadomości – myślę, że pokazałem dlaczego. Niebezpośredniość metody heterofenomenologicznej pozwala uniknąć nieuzasadnionego obowiązku „identyfikowania" czy „redukowania" (domniemanych) bytów zamieszkujących ontologię osób badanych. Czy antropologowie *identyfikują* Feenomana z poznanym chłopakiem, który dokonuje wszystkich dobrych czynów w dżungli, czy są „eliminacjonistami", jeśli chodzi o Feenomana? Gdyby dobrze wykonali *swoją* pracę, jedyna nierozstrzygnięta kwestia wymagałaby decyzji z zakresu polityki dyplomatycznej, nie doktryny naukowej czy filozoficznej. W pewnym sensie można by powiedzieć, że moja teoria identyfikuje świadome przeżycia ze zdarzeniami w mózgu niosącymi informacje – skoro dzieje się tylko to, wiele zdarzeń w mózgu jest niezwykle podobnych do mieszkańców światów heterofenomenologicznych podmiotów. Jednak inne właściwości elementów heterofenomenologicznych mogą zostać uznane za „esencjalne" – na przykład pozycja, jaką zajmują te elementy w subiektywnych sekwencjach czasowych. Wówczas *nie mogłyby* bez pogwałcenia prawa Leibniza zostać utożsamione z dostępnymi zdarzeniami mózgowymi, które mogą przebiegać w innej kolejności.

Kwestia dotycząca tego, czy traktować część heterofenomenologicznego świata podmiotu jako użyteczną fikcję, czy może raczej jako naciąganą prawdę, nie zawsze jest sprawą zasługującą na wiele uwagi. Czy wyobrażenia umysłowe są prawdziwe? Istnieją prawdziwe struktury danych w ludzkich mózgach, które są właściwie wyobrażeniami – czy *to* są wyobrażenia umysłowe, o które pytamy? Jeśli tak, to owszem; jeśli nie, to nie. Czy *qualia* są definiowalne funkcjonalnie? Nie, ponieważ nie istnieją właściwości takie jak *qualia*. Lub też nie, gdyż *qualia* są dyspozycyjnymi właściwościami mózgów, które nie są ściśle definiowalne w kategoriach *funkcjonalnych*. Lub tak, bo jeśli naprawdę zrozumielibyśmy wszystko na temat funkcjonowania układu nerwowego, to zrozumielibyśmy wszystko na temat właściwości, o których ludzie rzeczywiście mówią, gdy twierdzą, że mówią o swoich *qualiach*.

Czy jestem więc funkcjonalistą? Tak i nie. Nie jestem funkcjonalistą maszyny Turinga, ale też wątpię, aby ktokolwiek kiedykolwiek nim był, a szkoda, bo musi się marnować aż tyle dobrych kontrargumentów. Jestem oczywiście rodzajem „teleofunkcjonalisty", być może pierwszym teleofunkcjonalistą (w *Content and Consciousness*), ale, jak cały czas powtarzam

i podkreślam w dyskusji o ewolucji i *qualiach*, nie popełniam błędu próby definiowania wszystkich najistotniejszych różnic umysłowych w kategoriach *funkcji* biologicznych. Oznaczałoby to rażąco błędne odczytywanie Darwina.

Czy jestem instrumentalistą? Myślę, że w artykule *Rzeczywiste wzorce* (1991a/2008) pokazałem, dlaczego jest to kiepsko postawione pytanie. Czy ból jest prawdziwy? Jest tak prawdziwy jak fryzury, dolary, szanse, ludzie i środki ciężkości, ale jak bardzo wszystko to jest prawdziwe? Te dychotomiczne pytania wynikają z zapotrzebowania na wypełnienie powyższej, skwantyfikowanej formuły, a niektórzy filozofowie uważają, że teorię umysłu tworzy się przez obmyślanie kuloodpornego sądu tego rodzaju, a następnie jego bronienie. Zwykły sąd logiczny nie jest teorią, jest sloganem; a niektórzy filozofowie nie teoretyzują, tylko udoskonalają slogany. *W jakim celu* to robią? Jaki mętlik zostałby uporządkowany, jakie byłyby postępy w otwieraniu nowych perspektyw, gdyby coś takiego się udało? Czy naprawdę potrzebujesz czegoś do wydrukowania na koszulce? Niektórzy cyzelatorzy sloganów są w tym niezwykle dobrzy, ale jak mądrze powiedział kiedyś psycholog Donald Hebb, „jeśli nie warto tego robić, nie warto tego robić dobrze".

Nie zamierzam twierdzić, że ostrożna definicja i jej krytyka przez kontrprzykłady nigdy nie są cennymi ćwiczeniami. Weźmy przykład definicji barwy. Ostatnie analizy i próby jej zdefiniowania przez filozofów były pouczające. Udało im się wyjaśnić to pojęcie i odeprzeć prawdziwe nieporozumienia. Biorąc pod uwagę troskę, jaką przejawiają w ostatnim czasie filozofowie starający się precyzyjnie zdefiniować pojęcie barwy, pogląd, że barwy są właściwościami odbijania światła od powierzchni przedmiotów lub od przezroczystych obiektów, jest skandalicznym uproszczeniem. Jakie dokładnie właściwości odbijania? Myślę, że wyjaśniłem, dlaczego próba odpowiedzi na to pytanie byłaby stratą czasu; jedyna naprawdę precyzyjna odpowiedź nie mogłaby być zwięzła, z dobrze rozumianych przez nas powodów. Oznacza to, że trudno spotkać „niekolistą" definicję. I co z tego? Czy naprawdę uważam, że to proste posunięcie może obalić stanowiska moich przeciwników? (Poza wcześniej cytowanymi autorami wspomniałbym też o Strawsonie 1989 oraz Boghossianie i Vellemanie 1989, 1991). Tak, ale ta historia jest długa, więc przerzucę piłkę na ich stronę.

Czy twoje stanowisko nie jest koniec końców odmianą weryfikacjonizmu?

W ostatnim czasie filozofowie zdołali przekonać siebie samych – oraz wiele niewinnych osób postronnych – że weryfikacjonizm *zawsze* jest grzechem. Pod wpływem na przykład Searle'a i Putnama, neuronaukowiec Gerald Edelman pośpiesznie wycofuje się ze stanowiska bliskiego weryfikacjonizmowi: „Nieobecność dowodu samoświadomości u zwierząt, z wyłączeniem szympansów, nie pozwala nam na uznawanie, że nie posiadają samoświadomości" (Edelman 1989, s. 280). Fuj! Odwagi! Z pewnością nie tylko możemy uznawać, że jej nie mają, ale możemy zbadać racje stojące za tym uznaniem, a jeśli znajdziemy silne, pozytywne powody do tego, aby mu zaprzeczyć, powinniśmy to zrobić. Czas, aby wahadło wychyliło się w drugą stronę. W komentarzu do mojej wcześniejszej krytyki Nagela (Dennett 1982a) Richard Rorty stwierdził kiedyś:

Dennett uważa, że można być sceptycznym wobec tezy Nagela na temat bogatego fenomenologicznie, wewnętrznego życia nietoperzy „bez równoczesnego stawania się przaśnym weryfikacjonistą". A ja nie. Myślę, że sceptycyzm wobec intuicji, o jakich mówią Nagel i Searle, jest przekonujący, tylko jeśli jest oparty na ogólnych względach metodologicznych dotyczących statusu intuicji. Ogólnym zarzutem weryfikacjonisty do realistów jest to, że twierdzi on, że istnieją różnice (na przykład między nietoperzami z życiem prywatnym i bez, psami z wewnętrzną intencjonalnością i bez), które nie mają znaczenia: to, że jego intuicje nie mogą zostać włączone w schemat wyjaśniający, ponieważ są „kołami, które można obracać bez

poruszania czegoś innego" [Wittgenstein 1953/2000, I, par. 271]. Wydaje mi się to dobrym zarzutem, a zarazem jedynym, który musimy wysunąć. [Rorty 1982a, s. 342–343; zob. też Rorty 1982b]

Zgodziłem się, lecz zaproponowałem delikatną (o 0,742) zmianę twierdzenia: „z dopingującym mi profesorem Rortym..., jestem gotów wyjść z szafy jako pewnego rodzaju weryfikacjonista, ale, proszę, oby nie przaśny; bądźmy wykwintnymi weryfikacjonistami" (Dennett 1982b, s. 355). Ta książka obiera ten kurs, przekonując, że jeśli nie będziemy wykwintnymi weryfikacjonistami, to zaczniemy w końcu tolerować wszelki nonsens: epifenomenalizm, zombi, nierozróżnialne odwrócone spektra, świadome pluszaki, świadome siebie pająki.

Najbardziej problematyczna kwestia dla tego rodzaju weryfikacjonizmu, jaki wspieram, pojawia się w rozdziale 5, w argumencie mającym na celu pokazanie, że skoro nie ma i *może nie być* dowodu na potwierdzenie modelu świadomości orwellowskiej lub stalinowskiej, to pozbywamy się tego dylematu. Standardowa reakcja na to weryfikacjonistyczne stwierdzenie jest taka, że z góry przesądzam bieg nauki; skąd wiem, że nowe odkrycia w dziedzinie neuronauki nie *ujawnią* nowych podstaw do poczynienia takiego rozróżnienia? Odpowiedź – nieczęsto dziś słyszana – jest prosta: w przypadku pewnych pojęć (nie wszystkich, ale niektórych) możemy być pewni, że wiemy wystarczająco dużo, a *cokolwiek* pojawi się w przyszłości w nauce, nie otworzy to takich możliwości. Rozważmy na przykład hipotezę, że wszechświat jest w pozycji „normalnej", i jej zaprzeczenie, że wszechświat jest do góry nogami. Czy są to dobre hipotezy? Czy jest lub mógłby to być problem? Czy jest weryfikacjonistycznym błędem uważać, że bez względu na to, jakie rewolucje nastąpią w kosmologii, nie zmienią tej „dysputy" na spór, który może być rozstrzygnięty przez fakty empiryczne?

Ale jesteś behawiorystą pewnego rodzaju, prawda?

To pytanie zostało już kiedyś zadane i chętnie wesprę odpowiedź, którą dał na nie Wittgenstein (1953/2000).

307. „Czy nie jesteś jednak zamaskowanym behawiorystą? Czy nie utrzymujesz w gruncie rzeczy, że poza ludzkim zachowaniem wszystko jest fikcją?" – Jeśli mówię tu o fikcji, to o fikcji gramatycznej.

308. W jaki sposób pojawia się filozoficzny problem zjawisk i stanów psychicznych oraz behawioryzmu? – Pierwszy krok jest całkiem niepozorny. Mówi się o stanach i zjawiskach, pozostawiając kwestię ich natury otwartą! Może kiedyś dowiemy się o nich więcej – myślimy sobie. Ale właśnie przez to związaliśmy się z określonym ujęciem sprawy. Mamy bowiem określone pojęcie o tym, *co* znaczy: poznać bliżej pewne zjawisko. (Decydujący krok w sztuczce kuglarskiej został zrobiony, a właśnie on zdawał się całkiem niewinny.) – I oto rozpada się porównanie, które miałoby uczynić nasze myśli zrozumiałymi. Trzeba więc zakwestionować ów niezrozumiany jeszcze proces w niezbadanym jeszcze medium. A wtedy wydaje się, że zakwestionowaliśmy zjawiska psychiczne. A przecież nie chcemy ich bynajmniej kwestionować!

Kilku filozofów postrzega to, co robię, jako rodzaj powtórzenia ataku Wittgensteina na „przedmioty" świadomego przeżycia. I właśnie tak jest. Jak wyjaśnia paragraf 308, jeśli chcemy uniknąć kuglarskiej sztuczki, musimy najpierw zrozumieć „naturę" psychicznych zjawisk i stanów. Dlatego potrzebowałem dziewięciu długich rozdziałów, aby dojść do momentu, w którym mogłem rozprawić się z problemami w ich typowo filozoficznym przebraniu – to znaczy w ich błędnym przebraniu. Mój dług wobec Wittgensteina jest ogromny i dawny. Gdy byłem studentem, był moim bohaterem, więc poszedłem na studia doktoranckie do Oksfordu, bo wydawało się, że jest tam bohaterem wszystkich. Kiedy zdałem sobie sprawę z tego, jak mało rozumieli go (według mnie) inni doktoranci, porzuciłem chęć „bycia" wittgensteinistą i po prostu

wykorzystałem w mojej pracy to, czego nauczyłem się z *Dociekań*.

Załącznik B (dla naukowców)

Filozofów często oskarża się o upodobanie do psychologii kanapowej (czy też neuronauki, fizyki itd.) i jest wiele wstydliwych opowieści o filozofach, których pewne aprioryczne deklaracje zostały następnie obalone w laboratorium. Jedna z racjonalnych reakcji na to znane ryzyko to ostrożne wycofanie się filozofa w takie obszary pojęciowe, gdzie jest małe bądź żadne ryzyko powiedzenia czegoś, co mogłoby zostać obalone (lub potwierdzone) przez empiryczne odkrycie. Kolejną racjonalną reakcją jest badanie, w zaciszu własnego gabinetu, najlepszych wyników laboratoryjnych, największych prac empirycznie ugruntowanych teoretyków, a następnie przejście do swojej filozofii i próba wyjaśnienia pojęciowych przeszkód, a nawet okazjonalne wystawianie się w taki czy inny sposób na ryzyko w celu jasnego wskazania następstw pewnych szczególnych idei teoretycznych. Jeśli chodzi o kwestie pojęciowe, naukowcy nie są bardziej odporni na zamęt niż amatorzy. Naukowcy spędzają w końcu sporo swojego czasu na kanapach, próbując się zorientować, jak interpretować rezultaty różnych eksperymentów, a to, co wówczas robią, niepostrzeżenie łączy się z tym, co robią filozofowie. Sprawa ryzykowna, choć pobudzająca.

Oto zatem kilka nie całkiem jeszcze rozwiniętych pomysłów na doświadczenia, mające przetestować następstwa przedstawionego przeze mnie modelu świadomości, które nie przebiły się przez ogień moich cierpliwych krytyków albo zostały uznane przez nich za już przeprowadzone. (Moje wyniki w tej drugiej kategorii są wystarczająco dobre, by zachęcić mnie do dalszego uporu). Jako filozof staram się, aby mój model był jak najogólniejszy i maksymalnie neutralny, jeśli więc mi się to udało, te eksperymenty powinny jedynie pomóc ustalić, *w jakim stopniu* jakaś wersja mojego modelu jest potwierdzona; jeśli model byłby zupełnie błędny, zostałby całkowicie obalony, a ja zawstydzony.

Czas i umiejscowienie w czasie

Jeśli subiektywna sekwencja jest wynikiem interpretacji, a nie bezpośrednio odwzorowuje prawdziwy ciąg zdarzeń, możliwe powinno być wytworzenie silnych, interpretacyjnych efektów różnego rodzaju, niezależnych od rzeczywistego umiejscowienia w czasie.

1. *Pająk chodzi*: delikatne, kolejne dotknięcia, podobne do „skórnego królika", ale mające wytworzyć iluzoryczne oceny *kierunku*. Prostym przypadkiem byłyby dwa dotknięcia rozdzielone w przestrzeni i czasie, w podobnym zakresie jak zjawisko wizualnego phi, a zadaniem byłaby ocena kierunku „chodzenia" (co jest logicznie równoważne *sekwencji*, ale fenomenologicznie sądem bardziej „bezpośrednim"). Moje przypuszczenie: standardowe efekty zjawiska phi zależą od długości przerwy między bodźcami, a większa wyrazistość powstanie na powierzchniach o wysokiej rozdzielczości, jak koniuszek palca czy usta.

Poprośmy jednak badanego, aby ustawił prawy i lewy palec wskazujący obok siebie, i doprowadźmy pierwszy bodziec do jednego koniuszka, a drugi do drugiego. Powinno nastąpić o wiele gorsze rozpoznanie kierunku ze względu na wymóg, że porównania muszą być bilateralne. Następnie dodajmy „pomoc" wizualną; pozwólmy badanemu obserwować stymulację palca, ale niech będzie to błędny obraz: przygotujmy urządzenia tak, aby dorozumiany *kierunek wizualny* był kierunkiem przeciwnym do kierunku wyznaczonego przez rzeczywistą sekwencję dotknięć. Moje przypuszczenie: badany będzie w sposób pewny wyrażał błędne oceny, odrzucając czy usuwając rzeczywistą sekwencję informacji dostępnych poprzez receptory dotykowe. Jeśli efekt będzie bardzo silny, może nawet odrzucić ocenę unilateralną czy

pochodzącą z jednego palca, która była bardzo trafna pod nieobecność informacji wizualnych.

2. *Odwrócenie filmu*: osoby badane proszone są o rozróżnienie krótkich „ujęć" kinowych czy filmowych, a niektóre z nich zostały odwrócone lub pojawiają się w nich zakłócenia bądź anomalne sekwencje. Montażyści mają swoje sztuczki oraz mnóstwo wiedzy tajemnej, związanej z niepoprawnym ustawianiem kolejności kadrów filmu. Czasem celowo łączą ze scenami kadry w nieodpowiedniej kolejności, aby wytworzyć efekty specjalne – na przykład, aby wyostrzyć lęk czy szok w scenach horroru. Niektóre wydarzenia w sposób naturalny są silnie uporządkowane; wszyscy cieszyliśmy się, widząc film, w którym nurek wyłania się, poczynając od stóp, z plusku w basenie i skacze, lekki i suchy, na trampolinę. Inne wydarzenia są niezauważalnie możliwe do odwrócenia – na przykład trzepocząca flaga – podczas gdy inne są pośrednie; może nie być tak łatwe stwierdzenie, czy odbijająca się piłka poruszała się do przodu, czy do tyłu. Moje przypuszczenie: ludzie nie będą sobie dobrze dawać radę z rozróżnianiem odwróceń, w których nie ma żadnej interpretacyjnej stronniczości – w których sama sekwencja musi zostać wykryta i zapamiętana. Na przykład, gdy ciągłość ruchu oraz rozmiar i kształt różnic są w miarę niezmienne, badani prawdopodobnie o wiele gorzej będą sobie dawać radę z rozróżnianiem (ponowną identyfikacją) sekwencji niemających stronniczej interpretacji kierunkowej oraz z odróżnianiem ich od ich odwróceń i innych przekształceń. (Eksperymenty z rozróżnianiem melodii byłyby odpowiednikiem słuchowym).

3. *Pisanie na stopie*: eksperyment zaprojektowany, aby zdezorganizować oceny oparte na interpretacji „czasów przybycia" do „centralnej dostępności". Przypuśćmy, że bierzesz ołówek i piszesz jakieś litery na boku swojej bosej stopy bez obserwowania tego, co robisz. Sygnały z receptorów dotykowych w twojej stopie „potwierdziłyby", że czynności intencjonalnego pisania zostały odpowiednio zrealizowane przez ołówek w twojej ręce. Teraz dodaj widzenie pośrednie, monitor telewizyjny, pokazujący, jak piszesz ręką na stopie, ale z kamerą umieszczoną w taki sposób, że koniuszek ołówka na stopie jest zasłonięty przez rękę trzymającą ołówek. Te sygnały wizualne dodałyby kolejne potwierdzenie wykonania twoich intencjonalnych czynności. Teraz jednak włóż do odtwarzacza taśmę z opóźnieniem (jedna czy dwie klatki, każda z nich długości 33 milisekund), aby potwierdzenie wizualne zawsze było odrobinę, ale stale, opóźnione. Przypuszczam, że badani szybko by się do tego dostosowali. (Mam taką nadzieję, gdyż kolejny krok jest bardzo ciekawy). Gdy już przyzwyczaili się do tej sytuacji, to gdyby nagle usunąć opóźnienie, zinterpretowaliby rezultat jako *poczucie, że ołówek się zgina*, ponieważ percepcja trajektorii jego koniuszka byłaby opóźniona względem informacji wizualnych, jak gdyby ciągnął się on na moment przed jego spodziewaną trajektorią.

4. *Dopasowywanie opóźnienia na karuzeli Greya Waltera*: dalszy eksperyment mierzący długość opóźnienia wymaganego, aby wyeliminować efekt „rzutnika prekognitywnego". Przypuszczam, że długość będzie znacznie mniejsza niż 300–500 milisekund, których można by się spodziewać przez rozszerzenie stalinowskiego modelu Libeta.

Modele i pandemonium doboru słów

Jak można pokazać, że „słowa chcą zostać wypowiedziane"? Czy dokonywanie przypadkowych odkryć może być kontrolowane eksperymentalnie? Dotychczasowe eksperymenty Levelta wykazały zaskakująco negatywne rezultaty (zob. przypis 77 na s. 174). Pewna ich odmiana, którą chciałbym zobaczyć, otworzyłaby możliwości „kreatywnego" użycia słów przez badanego, jednocześnie dyskretnie dostarczając różnych surowców do środowiska, aby włączył je on w swoje realizacje językowe. Na przykład badany mógłby być przygotowywany do eksperymentu w dwóch różnych, wstępnych otoczeniach, w których różne

słowa: zaskakujące, żywe, odrobinę nowe czy nie na miejscu zostałyby „przypadkiem" pozostawione wokół niego (na plakatach na ścianie, w instrukcjach dla niego itp.); następnie badanemu dano by szansę, aby wyraził się na tematy, w których te docelowe wyrażenia nie miałyby dużej szansy się pojawić, a w taki sposób ugruntowanie we wstępie pokazałoby, że docelowe wyrażenia zostały „włączone" i się czaiły, szukając możliwości bycia użytymi. Nieodnalezienie żadnego efektu wsparłoby model Levelta; odnalezienie dużego efektu (zwłaszcza gdyby wykorzystane zostały „wymuszone" możliwości) wsparłoby model pandemonium.

Eksperymenty z wykorzystaniem okulografu

1. *„Ślepowidzenie" u normalnych badanych*: eksperymenty z normalnymi osobami badanymi wykorzystujące okulograf pokazały, że gdy bodziec na obrzeżu dołka środkowego jest zmieniony podczas sakkady, badani tego nie zauważają (nie relacjonują żadnego odczucia zmiany), ale pojawiają się efekty rozszerzenia – czas identyfikacji drugiego bodźca jest skrócony lub nie zależy od informacji zebranych z oryginalnego bodźca na obrzeżu dołka środkowego. Jeśli badani w takich warunkach dokonują wymuszonego odgadnięcia dotyczącego tego, czy bodziec został zmieniony (lub czy pierwotny bodziec był, powiedzmy, wielką czy małą literą), to czy ilość prawidłowych odgadnięć będzie większa niż przypadkowa? Przypuszczam, że tak, z wielu interesujących powodów, ale nie większa niż w najlepszych przypadkach ślepowidzenia.

2. *Eksperymenty „tapetowe"*: używając okulografu i zmieniając zarówno spore, jak i niewielkie właściwości powtarzających się pól wzorów na „tapecie" na obrzeżach dołka środkowego podczas sakkad, ustal, co mogłoby oddalić wniosek o „obrazach wielu Marilyn". (Zaskoczyły mnie najnowsze rezultaty Ramachandrana i Gregory'ego, więc zaryzykuję i powiem, że prawdopodobnie *nie* zaistnieją wykrywalne *stopniowe* efekty, chociaż na poziomach, na których badani zauważą zmiany, mogą oni również odebrać je jako dziwne, złudne ruchy).

3. *Barwna szachownica*: eksperyment zaprojektowany, aby pokazać, jak niewiele jest w „obszarze pola widzenia". Badani dostają zadanie identyfikacji bądź reidentyfikacji wymagającej wielu sakkad po poruszającej się scenie: oglądają animowane, czarno-białe postacie pokazane na tle losowo pokolorowanej szachownicy. Pola na niej są dosyć duże – na przykład ekran jest podzielony na sieć 12×18 barwnych kwadratów losowo wypełnionych różnymi barwami. (Barwy wybrane są losowo, aby ich schemat nie miał żadnego znaczenia dla zadania wzrokowego nałożonego na to tło). Pomiędzy kwadratami powinny być różnice jasności, aby nie pojawił się efekt Liebmanna, a dla każdego kwadratu powinna być przygotowana *barwa konkurencyjna o tej samej jasności*: taka, która w przypadku zamiany z barwą aktualnie wypełniającą kwadrat nie stworzy radykalnie innych krawędzi jasności na brzegach (wszystko to, aby uciszyć detektory krawędzi i jasności). Następnie załóżmy, że podczas sakkad barwy na szachownicy zostają zmienione; obserwatorzy zauważyliby, jak jeden lub więcej kwadratów zmienia barwę kilka razy na sekundę. Moje przypuszczenie: zaistnieją warunki, w których badani będą całkowicie nieświadomi faktu, że spora część „tła" nagle zmienia barwę. Dlaczego? Ponieważ system wzrokowy obrzeża dołka środkowego jest przede wszystkim systemem alarmowym, złożonym z wartowników zaprogramowanych tak, aby wywołać sakkady, gdy zostaje zauważona zmiana; taki system nie zawracałby sobie głowy śledzeniem nieistotnych barw pomiędzy fiksacjami, a więc nie pozostałoby mu nic, z czym mógłby porównać nową barwę. (Zależy to oczywiście od tego, jak „szybki jest film" w obszarach reagujących na kolor na obrzeżach dołka środkowego; może zaistnieć długaśny okres refrakcji, który zlikwiduje

przewidywany przeze mnie efekt).

Posłowie:

Jak nie czytać Dennetta?

Książka *Świadomość* ukazała się po angielsku w roku 1991, czyli ćwierć wieku temu. Mogłoby się zdawać, że nie ma sensu jej tłumaczyć i publikować w Polsce, bo badania nad świadomością poszły dalej – autor przecież nie jest jasnowidzem i nie był w stanie przewidzieć wszystkiego, co odkryto później. Tak też sądził wydawca książki *Słodkie sny* (Dennett 2007), której przekładu dokonałem w roku 2007 (w oryginale wyszła drukiem dwa lata wcześniej). Poprzedziłem ją wówczas wstępem, gdyż bez uwzględnienia tego, co Daniel Dennett napisał w *Świadomości*, *Słodkie sny* są mało zrozumiałe. W istocie jednak to *Świadomość* powinna była się ukazać wcześniej.

Pomimo upływu dwudziestu pięciu lat pozostaje niesłychanie wpływowa jako punkt odniesienia – zarówno negatywny, jak i pozytywny. A jednocześnie jest przedmiotem nieustannych przeinaczeń i krytyki. W *Słodkich snach* Dennett starał się odpowiedzieć na chybione zarzuty, lecz legendy, którymi obrastają niektóre książki, czasem przesłaniają ich rzeczywistą treść. Nakreślę więc pewne mity, a potem pokrótce zarysuję, jak zmieniały się badania nad świadomością w ostatnim ćwierćwieczu. Na deser zostawię zaś pytanie, jak należałoby krytykować Dennetta, żeby dyskusja była naprawdę merytoryczna.

1. Krwiożerczy redukcjonizm

Najczęstszym zarzutem stawianym Dennettowi jest to, że eliminuje on świadomość. Angielski tytuł to *Consciousness Explained* – a więc dosłownie „Świadomość wyjaśniona", a może lepiej „Wyjaśnianie świadomości"; krytycy dodają słówko „away" – co miałoby znaczyć „Świadomość zdemaskowana jako pozór". Dennett bowiem krytykuje filozoficzne pojęcie *qualiów*. *Quale* (czytaj: kwale) to wyraz pochodzący z łaciny: to forma mianownika liczby pojedynczej rodzaju nijakiego wyrazu *qualis*, odpowiadającego dosyć ściśle polskiemu przymiotnikowi „jaki". Można od niego rozpoczynać pytania i zdania podrzędne. W języku angielskim brak słówka, które wyrażałoby różnicę między „jaki" i „który", więc na określenie *jakości*, oprócz wyrazu *quality*, używa się też wyrazu z łaciny *quale*, znaczącego wówczas: to, jakie coś jest. Ta wycieczka etymologiczna, może nieco nużąca, jest nam potrzebna, gdyż warto od razu zauważyć, że Dennett, skupiając się na pojęciu *quale*, bynajmniej nie krytykuje żadnego potocznego pojęcia świadomości. Wyraz ten nie występuje w zwykłej angielszczyźnie i wchodzi do użycia filozoficznego dopiero w XX wieku. Mówiąc krótko, Dennett uważa, że wprowadzony do angielszczyzny neologizm nie jest przydatny do opisu jakości przeżyć świadomych. Czemu więc oskarża się go o eliminowanie samej świadomości?

Powód jest prosty: wielu badaczy, zarówno filozofów, jak i kognitywistów, twierdzi, że najistotniejszym aspektem świadomości jest jej jakościowy charakter. Mylnie jednak sądzą, że Dennett uważa, iż nie przeżywamy swoiście wyglądu dojrzałego jabłka – w przeciwieństwie do tego, jak pachnie. Dennett krytykuje jedynie ideę, że te wyglądy, zapachy, kształty opisuje poprawnie pojęcie *quale*. Eliminuje więc pojęcie *quale*, ale nie to, co miało oznaczać. Zgodnie z tymi filozoficznymi teoriami *qualia* są jednocześnie własnościami wewnętrznymi, niewyrażalnymi językowo, dostępnymi wyłącznie dla jednej osoby, a wiedza na ich temat nie podlega jakiejkolwiek korekcie i jest bezpośrednio dostępna, bez żadnego wnioskowania. Dennett pokazuje, że przeżycie świadome nie ma takich własności i że często nasza wiedza na

temat własnych przeżyć podlega korektom i jest wyrażalna (bo potrafimy się dobrze komunikować). A jeśli tak, to nie są one dostępne tylko dla jednej osoby. Podstawowym jednak argumentem jest to, że gdyby takie *qualia* istniały, byłyby empirycznie niewykrywalne, a postulowanie empirycznie z zasady niewykrywalnych bytów w teorii empirycznej (a za taką teoria *qualiów* ma uchodzić) jest dla Dennetta niedopuszczalne.

W eseju *Quining Qualia* Dennett na wielu przykładach pokazywał, że takie pojęcie *qualiów* jest niespójne: nie można przypisywać im wszystkich wymienionych cech jednocześnie (Dennett 1988). Podkreślał na przykład, że nie sposób odróżnić, subiektywnie czy obiektywnie, czy zmieniają się *qualia*, czy nasze reakcje na *qualia*, a wówczas pojęcie *qualiów* – mające opisywać zjawisko empiryczne – nie ma empirycznych kryteriów stosowania. Oto i eksperyment myślowy Dennetta. Panowie Chase i Sanborn pracują w Maxwell House jako degustatorzy kawy. Pewnego dnia, po sześciu latach pracy Chase odzywa się do Sanborna: „Wiesz co, przez tyle lat wyszlachetnił mi się smak i już nie mogę znieść tej lury, którą tu serwują!". Na co Sanborn dziwi się i mówi: „Stary, wiesz, ja mam ten sam gust, ale przez te lata zużyły mi się chyba kubki smakowe, bo też nie mogę znieść tej kawy. Muszę się przekwalifikować".

Nie jesteśmy w stanie rozstrzygnąć, który ma rację: czy *qualia* smaku kawy pozostały niezmienne, a zmieniały się reakcje na nie? A może zmieniała się dostępność *qualiów* smaku kawy w przeżyciu, które towarzyszy łykowi kawy? Skoro z definicji kubki smakowe nie mogą mieć wpływu na to, jakie są *qualia* przeżycia świadomego, to dlaczego po zjedzeniu strasznie kwaśnych ogórków słodka papryka wydaje się słodsza? W jaki sposób na *qualia* działa słodycz papryki, skoro *qualia* muszą istnieć tylko ze względu na strukturę przeżycia? Czyżby świadome przeżycia na mocy jakiegoś nieznanego prawa musiały zawsze dopasowywać się do czysto funkcjonalnych i strukturalnych własności umysłu?

Można by postulować taki dualizm, ale nie posunie nas to naprzód w sporze między Sanbornem i Chase'em. Co gorsza, obaj nie mogą mieć, jak chciała tradycja kartezjańska, bezpośredniej i niepowątpiewalnej wiedzy na temat *qualiów*, gdyż porównując starsze *qualia* w pamięci, mogą się pomylić. Czy najczęściej nie musimy brać na wzór płytek terakotowych do sklepu, jeśli chcemy dokupić takie same?

W podobny sposób Dennett rozprawia się w tej książce z zombi. Zombi to w filozofii umysłu nasze ścisłe fizyczne odpowiedniki, które mają być jednak nieświadome. Mają one być pozbawione *qualiów*, lecz rozmawiać z nami i móc reagować tak samo jak my. Według argumentów zwolenników dualizmu istnienie zombi jest pojęciowo możliwe (Chalmers 2010). Z tego wywodzą oni wniosek, że fizykalizm jest fałszywy, gdyż w przeciwnym razie zombi byłyby niemożliwe pojęciowo, czyli byłyby sprzeczne. A tej sprzeczności przecież nie widać. Tymczasem, zdaniem Dennetta, zombi nie dałoby się odróżnić od nas, czyli teoria zombi jest oparta na zakładaniu istnienia niewykrywalnych z zasady empirycznie różnic. Twierdzi więc, wbrew wielu ostrożnym zwolennikom fizykalizmu, że zombi tylko pozornie można pomyśleć. W istocie jednak jest to pojęcie puste, a cała debata o zombi jest stratą czasu.

Mówiąc krótko, Dennett bynajmniej nie sądzi, że nie mamy przeżyć zmysłowych. Nie uważa też, jakoby świadomość nie istniała. Wręcz przeciwnie – jego teoria ma wyjaśnić właśnie świadomość, a raczej ma dać podwaliny pod pełne, poprawne metodologicznie wyjaśnienie. Dennett nie kryje bowiem, że kreśli jedynie wstępny szkic tego, co w przyszłości mogłoby się stać pełną teorią świadomości.

Warto w tym miejscu zauważyć, że Dennett jest w pewnym sensie redukcjonistą. Redukcjonizm to pogląd, zgodnie z którym pewne rzeczy – czy to teorie, pojęcia, czy też byty – redukują się do innych bądź do nich sprowadzają. Bez wchodzenia w szczegóły (w języku polskim piszą o sprawie wyczerpująco Katarzyna Paprzycka [2005] i Robert Poczobut [2009])

można powiedzieć, iż redukcjonista w sprawie świadomości uważa, że nie jest ona fundamentalnym bytem, lecz czymś, co sprowadza się do bytów fizycznych. Podobnie redukcjonista w sprawie błyskawic uważa, że są one widocznymi wyładowaniami elektrycznymi w atmosferze, nie zaś czymś, co przekracza naszą współczesną fizykę. Jednak i świadomość, i błyskawice okazują się czymś realnym po zabiegu redukcji. Jest to tak zwana redukcja zachowawcza (nieeliminacyjna).

W przeciwieństwie do niej redukcja eliminacyjna polega na ukazaniu pozorności pewnego bytu. I tak we współczesnej nauce wyeliminowano pojęcie „histeria", zastępując je pojęciami opisującymi różnego rodzaju zaburzenia psychiczne (jeśli rzeczywiście zachodziły; czasem bowiem histerię przypisywano po prostu kobietom walczącym o swoje prawa wyborcze). Tak więc Dennett podobnie chce usunąć „*quale*" z listy terminów rzetelnie opisujących świadomość, tak jak „histerię" usunięto ze słownika psychiatrii.

To jednak niejedyna redukcja eliminacyjna, jakiej dokonuje. Filozofowie uważają, że świadomość cechuje się swoistą jednością i że istnieje jeden strumień świadomości. Byłyby to więc cechy przeżycia świadomego, które należałoby wyjaśnić. Tymczasem Dennett wskazuje, że ta jedność może powstawać w wyniku nie tyle jednorodnego i prostego procesu, ile czegoś o naturze złożonej, a w dodatku przebiegającego w innej kolejności niż dostępna nam w strumieniu świadomości.

Świadomość, jak mówi hasłowo Dennett, jest to wirtualna maszyna von Neumanna, zaimplementowana na maszynie z przetwarzaniem równoległym na wielką skalę. Maszyna von Neumanna to tradycyjna architektura komputera, która wykonuje polecenia po kolei, tak jak się pojawiają. Takie przetwarzanie nazywa się w informatyce „szeregowym". Natomiast przetwarzanie równoległe danych stanowi klucz do wielu koncepcji Dennetta. U innych teoretyków furorę w latach 70. robiła „pudełkologia" (jak to określał Dennett): kolejne czarne skrzynki na drodze przetwarzania danych przez organizm. Dennett poszedł nieco inną drogą. Przyjmując istnienie odrębnych jednostek przetwarzania, homunkulusów, uznał, że bynajmniej nie są one elementem biurokratycznego, scentralizowanego systemu, lecz raczej modelu przypominającego pandemonium, chaotyczną walkę wszystkich ze wszystkimi. Nie istnieje realny strumień świadomości, przechodzący kolejno przez rozmaite moduły, wytwarzane jest raczej poczucie jednego strumienia świadomości, podczas gdy przetwarzanie ma charakter równoległy.

Te śmiałe hipotezy popiera Dennett materiałem empirycznym, zwłaszcza eksperymentami, w których zaburzony wydaje się czas postrzegania pewnych zjawisk. Weźmy przykład interesującego złudzenia optycznego. Położone w niewielkiej odległości od siebie dwie lampki, czerwona i zielona, zapalają się po kolei. Wszyscy ludzie obserwujący przełączenie lampek w odpowiednich warunkach zauważają płynne przejście zapalonego punktu z jednego miejsca w drugie, przy czym według obserwatora punkt zmienia kolor w połowie drogi. Wiadomo jednak, że druga lampka zapala się później, obserwacja zaś skłania do wniosku, iż zapaliła się, zanim fizycznie zaszło zapalenie lampki. Jak to się dzieje? Czyżby mózg dostarczał nam informacje o przyszłych wydarzeniach? A może dokonuje się jakaś korekta *post factum*? Dennett rozważa dwa modele takiej korekty (orwellowski i stalinowski, jak sam je określa, nawiązując do dwóch różnych sposobów zafałszowywania historii). Stwierdza jednak, że sporu między oboma stanowiskami nie można rozstrzygnąć empirycznie: różnice między nimi są z zasady niewykrywalne. To stanowi argument przeciwko modelowi jednego strumienia świadomości. Ale jeśli pozbędziemy się kartezjańskiego przekonania, że świadomość polega na obserwacji przedstawień w teatrze świadomości, to nie będziemy musieli postulować wątpliwych korekt *post factum*. Zjawisko da się wszak wyjaśnić przez konkurencyjne przetwarzanie

informacji przez wiele elementów, którego wyniki nie zawsze docierają do świadomości, bo często są odrzucane przed wykonaniem kolejnych operacji („wojna wszystkich elementów przetwarzających ze wszystkimi elementami przetwarzającymi").

I tu dochodzimy do sedna krytyki Dennetta: uważa on, że nie trzeba zakładać jednego „miejsca" w mózgu, gdzie wszystko się ze sobą łączy, występując na scenie przed... No właśnie, przed kim? Przed oczyma homunkulusa, który w swojej głowie ma kolejny teatrzyk, gdzie procesy poznawcze występują przed oczami jego homunkulusa, i tak w koło Macieju? Jedność świadomości to jedynie wytwór złożonych procesów równoległych.

Sprawę łatwo wyjaśnić na przykładzie podróży metrem (Akins 1996). W pewnym mieście dawno temu istniała jedna linia metra. Kiedy Wacław przyjeżdżał przed Ludwikiem, to znaczy, że wcześniej wsiadł. Ale od kiedy pojawiło się siedem linii i można zaplanować różne przesiadki, to Wacław, który jest bardzo sprytny, może lepiej zaplanować swoją podróż i dotrzeć przed Ludwikiem, mimo że ten ostatni wsiadł pierwszy do wagonika metra. Podobnie jest z procesami przetwarzania informacji – wiele z nich łączy się w wielu miejscach z innymi, stąd też nie ma żadnej gwarancji, że przebieg zdarzeń w otaczającym nas świecie będzie odwzorowany przez kolejność uświadamiania sobie ich przez nas. Może i dobrze, bo niektóre z powstających w ten sposób iluzji ułatwiają zbudowanie telewizorów i wyświetlanie nieruchomych obrazów, które odbieramy jako film. Muchy tak dobrze nie mają. Nie warto ich brać ze sobą do kina.

Model świadomości, jaki proponuje Dennett, określa on mianem „modelu wielokrotnych szkiców". To określenie nawiązujące do przeżyć autora tekstu, który krąży w wielu wersjach u wielu osób, a następnie wraca z wieloma różnymi poprawkami. Ich połączenie w sensowną całość bywa czasem trudne, bo wcześniejsze poprawki mogą być mniej istotne niż późniejsze... Równolegle pracujące szlaki przetwarzania tworzą cząstkowe szkice, które w różny sposób mogą składać się na powstające świadome przeżycia. W późniejszych pracach Dennett nieco zmienił język opisu tego modelu. Nie mówi już o szkicach tekstów, lecz o sławie mózgowej, która odbija się echem w różnych częściach układu poznawczego. Podkreśla też, że przy analizie zjawisk świadomych koniecznie trzeba brać pod uwagę konsekwencje: co dzieje się dalej? Co dzieje się po zarejestrowaniu przez oko bodźców fizycznych? Co dzieje się, gdy myśl staje się świadoma? Jak to się dzieje, że wypiera ją inna myśl w strumieniu świadomości? Tak czy inaczej – model ten pozostaje stosunkowo stabilny. Zastanówmy się więc, jak ma się do współczesnych badań nad świadomością.

2. Współczesne teorie świadomości

Koncepcja Dennetta jest zbliżona do koncepcji globalnej przestrzeni roboczej, rozwijanej od lat przez Bernarda Baarsa (1988; 1997). Zgodnie z tą ostatnią świadome stają się te stany przetwarzania informacji w systemie poznawczym, które pojawiają się w swoistej, globalnej przestrzeni roboczej. Warto zauważyć, że ta globalna przestrzeń nie jest teatrem kartezjańskim: globalna przestrzeń jest przestrzenią w sensie informatycznym. Idzie w rzeczywistości o globalną dostępność pewnych informacji dla różnych podsystemów poznawczych. Dzięki temu osoby badane mogą odpowiadać na pytania o stany swojej świadomości: są one dostępne dla procesów odpowiadających za introspekcję. Te dostępne globalne stany regulują także zachowania tych osób. Teoria ta jest rozwijana zarówno w ujęciu neuronaukowym, jak i w badaniach nad sztuczną inteligencją.

Teoria Baarsa tłumaczy między innymi ograniczoną pojemność świadomości (przez ograniczoną pojemność przestrzeni roboczej), jej sekwencyjną naturę, a także to, że zdarzenia

świadome mogą generować nieświadome procesy mózgowe. Baars podkreśla jednak, że globalna przestrzeń robocza jest blisko powiązana ze świadomymi przeżyciami, lecz samo znajdowanie się w globalnej przestrzeni nie jest równoznaczne z byciem przeżyciem świadomym. Innymi słowy, Baars nie ujmuje jakościowego charakteru zdarzeń świadomych, lecz stara się pokazać architekturę procesów przetwarzania informacji, które prowadzą do powstania między innymi świadomych przeżyć. Teoria ta jednak wyjaśnia, w jaki sposób stany przetwarzania informacji stają się stanami świadomymi, a więc pozwala rozróżniać stany nieświadome i świadome. Jest zatem teorią stanów świadomych, a nie świadomego podmiotu (i nie wyjaśnia na przykład różnicy między jawą a snem).

Poważną konkurencją wobec idei Dennetta i Baarsa jest teoria myśli wyższego rzędu (*Higher-Order Thought*, HOT) Davida Rosenthala (2005), rozwijana w Polsce w laboratorium Michała Wierzchonia (2013). Ona również ma na celu wyjaśnienie procesu uświadamiania, a jednocześnie swoistej struktury samoświadomości towarzyszącej świadomości (przynajmniej u człowieka). Zdaniem Rosenthala myśl staje się świadoma, gdy jest przedmiotem innej myśli. Same myśli wyższego rzędu rzadziej bywają uświadamiane, lecz są przyczyną uświadamiania innych. W pewnej mierze teoria ta przypomina więc tradycyjną koncepcję zmysłu wewnętrznego.

Na przecięciu informatyki, modelowania statystycznego i neuronauki znajduje się natomiast integracyjna teoria informacji. Jej twórcą jest Giulio Tononi (2004). W tym ujęciu świadomość cechuje się dwiema podstawowymi cechami: (1) ma naturę informacyjną; (2) jest silnie zintegrowana. Tononi opracował miarę złożoności integracyjnej dla sieci przesyłających informacje, która ma służyć do eksperymentalnego pomiaru stopnia świadomości. Miara ta nie jest na razie doskonała (zmienia się ona dosyć diametralnie w czasie dla tej samej sieci, a więc złożoność zależy silnie od stanu początkowego, w którym rozpoczyna się pomiar), jest także bardzo trudna do efektywnego zastosowania, lecz zainspirowała szeroko zakrojone badania nad znalezieniem innych, łatwiejszych w użyciu ujęć (do sprawy jeszcze wrócę). Do integracyjnych teorii, czyli opartych na pewnej mierze złożoności, należą też badania Christofa Kocha (2008). Świadomość ma być efektem integracji informacji w sieci nerwowej; neuronalne korelaty świadomości mają się zatem cechować swoistym stopniem zintegrowania (synchronizacji).

Warto zauważyć, że rola integracji i złożoności jest również bardzo istotna dla teorii globalnej przestrzeni roboczej – jednym z kryteriów wyróżniania tej przestrzeni ma być gęstość oddziaływań przyczynowych. Dotychczasowe propozycje obliczeniowych teorii świadomości mają charakter wstępny, lecz zauważalna jest pewna konwergencja różnych ujęć informatycznych; samo przejście na poziom ilościowy świadczy o dojrzewaniu teorii. Jak jednak zauważył Ray Jackendoff (1990: 17), sama złożoność może być co najwyżej wyznacznikiem powstawania stanów świadomych, nie zaś pełnym wyjaśnieniem, dlaczego one zaistniały.

Jackendoff podkreśla, że nie wystarczy ujęcie struktury procesów prowadzących do powstania świadomości; konieczne jest także zbadanie struktury nośników informacji, czyli formatu reprezentacji świadomych. Zdaniem Jackendoffa tylko informacje pośredniego szczebla mogą być treścią świadomości. Czyste dane percepcyjne docierające z organów zmysłowych nie są nam dostępne w świadomości (najniższy poziom reprezentacji); podobnie nie jesteśmy w stanie operować abstrakcyjnymi myślami w oderwaniu od modalności zmysłowych. Na przykład wyrazami operujemy zawsze w reprezentacji fonologicznej (lub graficznej, jeśli są to wyrazy języka tylko pisanego); nie są nam one dostępne w postaci czysto pojęciowej. Koncepcja Jackendoffa, choć powstawała nieco wcześniej od popularnych obecnie, jest bliska ujęciu Baarsa: bycie świadomym polega na znajdowaniu się w pamięci krótkoterminowej (STM), charakterystycznej dla danej modalności zmysłowej; reprezentacje z STM, na które skierowano uwagę, znajdują się w samym centrum świadomości. Ujęcie Jackendoffa rozwija obecnie Jesse

Prinz (2012), który w swoich badaniach akcentuje zmysłową genezę wielu procesów poznawczych.

Nieco inny charakter od poprzednich koncepcji, wyjaśniających głównie proces uświadamiania stanów świadomości, czy to pojedynczo, czy kolektywnie, ma teoria świadomości jako modelu świata (Johnson-Laird 1983; Metzinger 2003). Koncepcja ta nie tylko nawiązuje do faktu, że istnieje samoświadomość, co podkreślają teorie typu HOT, i że świadome informacje są globalnie dostępne oraz wykazują się daleko idącą integracją, ale że integracja ta służy do budowania modelu wykorzystywanego do działania w świecie. Jednym z tych modeli jest model jaźni, który wyjaśnia integrowanie się uświadamianych informacji i ich znaczenie dla działania. Podmiotowość związana ze świadomością wiąże się, zdaniem Metzingera, z istnieniem fenomenalnego modelu jaźni, będącego wewnętrzną i dynamiczną reprezentacją organizmu, która nie jest rozpoznawana jako reprezentacja (jest przezroczysta). Mówiąc w największym uproszczeniu i metaforycznie, jedną z kluczowych funkcji ludzkiej świadomości ma być ciągłe przedstawianie samej relacji jawienia się świata w postaci jego dynamicznego modelu. Będąc świadomymi, otwieramy przed sobą świat. Podobne, symulacyjne ujęcie świadomości prezentuje Antti Revonsuo (2006), który twierdzi, że świadomość jest rodzajem wirtualnej rzeczywistości. Ta wirtualna rzeczywistość pojawia się też we śnie, a jawa różni się tym, że następuje większa interakcja ze środowiskiem.

Mimo że przedstawione teorie mają charakter do pewnego stopnia konkurencyjny, warto zauważyć, iż w wielu miejscach są zgodne. Świadomość nie wymaga istnienia teatru kartezjańskiego, lecz jest kwestią integracji i złożonego przetwarzania informacji. Sam Dennett też patrzy przychylnie na rozwijającą się koncepcję Metzingera, gdyż ten ostatni nie zakłada automatycznie, że wszystkie tradycyjnie przypisywane świadomości cechy są realne, ale wszystkie chce wyjaśnić. Kontrowersyjne natomiast jest to, czy rzeczywiście świadomość wymaga, by istniała myśl, która ją uświadamia; gdy jednak weźmiemy pod uwagę, że w modelach psychologicznych uświadomienie to polega na istnieniu pewnych złożonych interakcji między myślą niższego i wyższego rzędu, dostrzeżemy, że nie są to idee bardzo od siebie odległe, choć oczywiście nie są tożsame.

Do pełnego zrozumienia świadomości nam daleko. Tu jedynie skrótowo opisałem to, czego Dennett nie opisuje szczegółowo, gdyż powstało już po opublikowaniu przezeń *Świadomości*. Warto jednak zauważyć, że jedyną koncepcją, w której nacisk kładzie się na jakości przeżyć, jest koncepcja Jackendoffa i Prinza. Ale tam też nie jest ona ujmowana w kategoriach *qualiów*, lecz raczej ucieleśnionego, zmysłowego odbioru rzeczywistości.

3. Nowe metody badań świadomości

Kiedy Dennett pisał swoją książkę, dostępne metody badawcze w neuronaukach były znacznie mniej wyrafinowane niż dziś. Przede wszystkim chodzi o różne metody obrazowania mózgu, w tym najpopularniejsze obecnie obrazowanie funkcjonalnym rezonansem magnetyczny (fMRI), który umożliwia pomiar przepływu utlenowanej krwi w strukturach mózgowych. To zaś, jak się uważa, skorelowane jest z aktywnością lub wstrzymaniem działania (inhibicją) danej okolicy mózgu. Obrazowanie fMRI ma jednak poważną wadę w badaniu świadomości: mianowicie ma zbyt słabą rozdzielczość czasową, gdyż opiera się na przepływie krwi, a ten nie jest błyskawiczny, lecz trwa znacznie dłużej niż przesyłanie sygnałów elektrycznych. Z tego względu istotne jest zastosowanie też innych metod pomiarowych, zwłaszcza elektroencefalografii, która ma mniejszą rozdzielczość przestrzenną (elektrody umieszcza się na czaszce badanego i zbierają one uśrednioną aktywność wielu obszarów mózgu), ale znacznie

lepszą – czasową. Prócz tego kluczowe znaczenie mają też metody inwazyjne – bezpośrednie pomiary za pomocą elektrod (u pacjentów przed operacjami neurologicznymi), a także rezultaty wypadków, w wyniku których dochodzi do ogniskowych uszkodzeń mózgu. Mniej dramatyczne, a lepiej kontrolowane i bezpieczne efekty ma TMS, czyli przezczaszkowa stymulacja magnetyczna. Za pomocą TMS można stymulować lub hamować działanie określonej okolicy mózgu u osób zdrowych, co pozwala sprawdzać nie tylko, czy zachodzi systematyczna korelacja między aktywnością poznawczą (w szczególności: świadomą) a aktywnością danej okolicy mózgu, lecz również to, czy związek ma naturę przyczynową, gdyż możemy dokonywać eksperymentalnych interwencji w ten proces.

Wszystkie te metody są w tej książce Dennetta jeszcze praktycznie nieobecne (pojawiają się wszakże w jego późniejszych tekstach, w tym w *Słodkich snach*). Tymczasem stały się one podstawą pewnego programu badawczego, pokrewnego heterofenomenologii, którą opisuje w tej książce. Przypomnijmy, na czym on polega. Przedstawiając heterofenomenologię, Dennett chciał obalić trzy błędne poglądy na metodologię prowadzenia badań nad świadomością:

(1) zdarzenia świadome nie istnieją;

(2) zdarzenia świadome istnieją, lecz są epifenomenami;

(3) zdarzenia świadome istnieją, lecz nie może ich badać nauka.

Jego zdaniem heterofenomenolog bada właśnie zdarzenia świadome i to nie jako epifenomeny, czyli uznaje, że są one istotne przyczynowo. W metodzie tej – jako metodzie badawczej – nie rozstrzyga się, czym jest świadomość ani przeżycia świadome, w tym sensie, że nie zawiera ona wstępnie opracowanego katalogu charakterystyk wszystkich przedmiotów umysłowych (na przykład czym różnią się doznania od halucynacji i czy oba są procesami jednego rodzaju) i tak dalej. Te kategorie być może zostaną dopiero opracowane w wyniku empirycznej pracy heterofenomenologów.

Heterofenomenologia stanowi po prostu ogólniejszy opis metody introspekcyjnej jako takiej, związany także z faktem, że w potocznej praktyce tak rozumiemy innych ludzi: pytając ich, co myślą. Metoda polega więc na badaniu językowych wypowiedzi osób relacjonujących swe przeżycia. Owe wypowiedzi traktuje się jako teksty. Osoby badane wypowiadają się swobodnie – nie znają intencji ani hipotez badacza, co mogłoby zakłócać ich sprawozdania słowne. Teksty powstające z wypowiedzi osób badanych traktuje się nie jako opisy rzeczywistości, lecz raczej tak jak fikcję w teorii literatury – czyli bez założenia, że opisywane obiekty istnieją czy też nie istnieją. Teoretyka literatury przy analizie powieści nie interesuje, czy świat przedstawiony odpowiada rzeczywistości, czy też nie. Podobnie ma postępować heterofenomenolog, a tekst sprawozdania introspekcyjnego konstytuuje intersubiektywnie dostępny świat heterofenomenologiczny.

Za pewne rozwinięcie heterofenomenologii może uchodzić współczesne badanie tzw. neuronalnych korelatów świadomych przeżyć (Metzinger 2000), o którym oczywiście Dennett nie mógł jeszcze ćwierć wieku temu wiedzieć. Warto zwrócić uwagę, że poszukiwania takich korelatów można prowadzić przy założeniu dualizmu, zgodnie z którym świadome przeżycia jedynie towarzyszą zjawiskom mózgowym. Założenie dualizmu zostaje jednak podważone, gdy tylko zastosujemy stymulację przezczaszkową; program korelacyjny nie jest więc ostatnim słowem w badaniach nad świadomością. Agnostycyzm – postawa głosząca, że nigdy nie będziemy wiedzieć, jakie stany mózgu wiążą się z jakimi procesami świadomości – wydaje się dziś zatem stosunkowo słabo uzasadniony, chociażby postępy w badaniach empirycznych były

powolne.

Jakie wyniki uzyskano w tych badaniach? Co najmniej dwa wydają się bodaj najbardziej spektakularne. Pierwszy wiąże się z tzw. zespołem zamknięcia – stanem pełnej przytomności i świadomości pacjenta w pełnym paraliżu. Taki pacjent nie może się komunikować z otoczeniem (zamknięcie jest stopniowalne, tak jak paraliż – relacja z częściowego zamknięcia znajduje się w zekranizowanej potem książce Jeana-Dominique'a Bauby'ego [2008]). Otóż badania z wykorzystaniem metod neuroobrazowania pozwalają coraz lepiej zdiagnozować takich pacjentów, a także budować urządzenia umożliwiające im komunikowanie się z otoczeniem (Górska i in. 2014). Są to różnego rodzaju interfejsy typu mózg-komputer. Wcześniej do diagnozy takich pacjentów stosowano bardzo proste metody, mające ułatwić stwierdzenie, czy znajdują się w śpiączce. Jednak w przypadku pełnego paraliżu nie ma możliwości zdiagnozowania u nich stanu minimalnej świadomości, gdyż na przykład wymagałoby to otwierania oczu przez pacjenta (spontanicznie lub w reakcji na bodziec bólowy czy polecenie).

Drugim ciekawym wynikiem jest uzyskanie możliwości pełnego skorelowania procesów mózgowych z obrazem widzianym przez osoby badane (Nishimoto i in. 2011). Eksperyment polegał na pokazywaniu badanym obrazów, a w trakcie badania za pomocą algorytmów uczenia maszynowego, czyli analizy statystycznych prawidłowości, udało się wychwycić zależność między aktywnością mózgu a konkretnym widzianym obrazem. Następnie dzięki danym z obrazowania mózgu można było zrekonstruować widziany przez badanego obraz. Chociaż ta metoda ma swoje ograniczenia – dotyczy tylko określonej grupy osób badanych („kod" mózgowy jest wysoce indywidualny), to pokazuje, iż tradycyjne poglądy na temat nieusuwalnej prywatności stanów świadomych można podważyć, a może nawet włożyć między bajki. Zauważmy, że ten eksperyment nie wymaga szczegółowych sprawozdań werbalnych, gdyż ze stanem mózgu korelowano bezpośrednio widziane obrazy (tylko co do nich badany zgadzał się, że je widzi). Jest zatem połączeniem metod heterofenomenologicznych i zwykłych badań z neuroobrazowania.

Kolejnym, obok łączenia sprawozdań słownych z aktywnością mózgu, ważnym elementem postępu w badaniach nad świadomością jest pojawianie się coraz doskonalszych, choć nadal dyskusyjnych, miar świadomości (Seth i in. 2008; Boly i in. 2013). Dlaczego to takie istotne? Otóż dopiero gdy zjawisko jest mierzalne, można mówić o precyzyjnych jego wyjaśnieniach, bo można sprawdzić, jak dokładnie zostało ono opisane czy przewidziane. Miary opierają się między innymi na stopniu integracji informacji w sieciach nerwowych oraz na wahaniach badanych co do podejmowanych przez nich decyzji i są stopniowo coraz bardziej udoskonalane – pierwsze wersje tych miar były niedoskonałe, gdyż, jak wspominałem, nie były praktycznie obliczalne. Dzięki możliwościom obrazowania i badania mózgu można było uzyskać spory postęp również w badaniu stanu mózgu w trakcie snu, ataku padaczki, śpiączce czy stanie wegetatywnym czy w znieczuleniu całkowitym i częściowym (notabene te badania prowadzono przy użyciu propofolu – tego samego leku, którego nadużył Michael Jackson i który wywołał jego śmierć). Co więcej, dysponując dobrą miarą, można nie tylko stwierdzać świadomość u osób sparaliżowanych, lecz także u zwierząt, które przecież nie mogą być podmiotami heterofenomenologicznymi w pełnym zakresie. Można je jednak wyuczyć odpowiednich zachowań, a przy tym stosować kryteria neuropsychologiczne, elektrofizjologiczne i psychofizyczne. Nie jesteśmy skazani na zgadywanie, czy (niektóre) koty świadomie są złośliwe, czy też tylko takie się wydają.

4. Jak czytać Dennetta, aby go krytykować?

Cóż zatem można, a nawet warto u Dennetta krytykować? Zarzut, że nie przewidział wszystkiego, byłby płytki. Zarzut, że odrzuca istnienie świadomości, byłby fałszywy.

Przede wszystkim warto zauważyć, że brak u niego pozytywnej propozycji dotyczącej jakościowych stanów świadomych. Odrzuca on pojęcie *quale*, lecz nie podaje żadnej pozytywnej charakterystyki jakościowego aspektu świadomości (w przeciwieństwie chociażby do Jackendoffa czy Prinza). Wręcz można mieć wrażenie, że zastępuje je pojęciem przekonania, a tymczasem pojęcie przekonania – w jego własnym ujęciu – jest jedynie przypisywane systemom intencjonalnym ze względu na prawidłowości w ich zachowaniach. Nie musi ono jednak opisywać nic realnego, jeśli chodzi o samą funkcjonalną architekturę poznawczą (Dennett 2003), a jedynie rzeczywiste wzorce w zachowaniu zewnętrznym (Dennett 2008). Co gorsza, pojęcie przekonania jest dyspozycyjne: jestem przekonany, że zebry nie są monitorami komputerowymi, ale nie cały czas o tym myślę. Myślę, że czytelniczki i czytelnicy też odkryją, że się z tym przekonaniem zgadzają, a więc mają – w sensie Dennetta – to przekonanie. Tymczasem jakości przeżyć świadomych są wyraziście ulotne, trudne do zapamiętania i odtworzenia, czym różnią się od zwykłych przekonań. Warto rozważyć, czy nie można mówić o jakościowych stanach świadomych takich organizmów jak ludzie czy inne ssaki w bardziej precyzyjny sposób niż przy użyciu dosyć ogólnikowego pojęcia „przekonanie".

Inną sprawą jest sam teatr kartezjański. Dennett przekonująco pokazuje, że nie musi on istnieć. Świadomość nie wymaga widza; nie oznacza to jednak, że niemożliwe jest istnienie organizmów, w których jednoznacznie da się wskazać granicę między świadomością a nieświadomością. Istnienie takiej klarownej granicy zdają się wskazywać koncepcje świadomości jako stanu umysłowego wyższego rzędu, co wcale nie świadczy o nich najgorzej. Istnienie teatru kartezjańskiego w takiej wersji nie może być bowiem logicznie wykluczone.

Kolejna sprawa wiąże się z samą heterofenomenologią. Z jednej strony jest to metoda introspekcyjna, a takie spotykają się współcześnie z licznymi problemami (Tyszka 1995; Hurlburt i Schwitzgebel 2007). Z drugiej strony konkuruje ona z metodami tzw. fenomenologii znaturalizowanej (Gallagher i Zahavi 2015). Ci ostatni krytykują heterofenomenologię, twierdząc, że jest zubożona względem oryginalnej fenomenologii Husserla i że można lepiej wyzyskać metody fenomenologiczne do badania świadomości. Do tej pory jednak nie ma zbyt wielu badań, które by rzeczywiście wskazały na rewolucyjne znaczenie samych metod znaturalizowanej fenomenologii i jej wyższość nad heterofenomenologią.

A najogólniej rzecz biorąc – propozycja Dennetta wymaga starannego rozważenia. Jak radzi sam Dennett, należy tak krytykować przeciwnika, by ten mógł siebie rozpoznać w krytykowanym stanowisku. Oto dokładny przepis:

Należy wyrazić stanowisko swojego przeciwnika tak jasno, dokładnie i sprawiedliwie, aby powiedział on: „Dziękuję, żałuję, że nie pomyślałem, aby ująć to w ten sposób".

Należy wyliczyć wszystkie punkty sporne (zwłaszcza jeśli nie są przedmiotem ogólnej czy powszechnej zgody).

Należy wspomnieć o wszystkim, czego można się nauczyć od przeciwnika.

Tylko wtedy ma się prawo wypowiedzieć choćby jedno słowo krytyki lub podjąć próbę odrzucenia krytykowanego stanowiska (Dennett 2015: 51).

Dopiero gdy najlepiej zrozumiana i poprawiona wersja jego stanowiska będzie nie do utrzymania, gdy rzeczywiście wykazane zostanie fiasko modelu wielokrotnych szkiców, będzie można ogłosić jej bankructwo. Do tej pory pozostanie ważnym stanowiskiem we współczesnej dyskusji nad świadomością[130].

Marcin Miłkowski

Instytut Filozofii i Socjologii PAN

Bibliografia

Akins K.A. 1989. *On Piranhas, Narcissism and Mental Representation: An Essay on Intentionality and Naturalism*, Ph.D. dissertation, Department of Philosophy, University of Michigan, Ann Arbor.

Akins K.A. 1990, *Science and Our Inner Lives: Birds of Prey, Bats, and the Common (Featherless) Biped*, [w:] M. Bekoff, D. Jamieson (red.), *Interpretation and Explanation in the Study of Animal Behavior*, vol. I, Westview, Boulder, CO, s. 414–427.

Akins K.A., Dennett D.C. 1986, *Who May I Say Is Calling?*, „Behavioral and Brain Sciences", *9*, s. 517–518.

Allman J., Meizin F., McGuinness E.L. 1985, *Direction and Velocity-Specific Responses from beyond the Classical Receptive Field in the Middle Temporal Visual Area*, „Perception", *14*, s. 105–126.

Allport A. 1988, *What Concept of Consciousness?*, [w:] Marcel, Bisiach (red.), 1988, s. 159–182.

Allport A. 1989, *Visual Attention*, [w:] M. Posner (red.), *Foundations of Cognitive Psychology*, MIT Press, Cambridge, s. 631–682.

Anderson J. 1983, *The Architecture of Cognition*, Harvard University Press, Cambridge, MA.

Anscombe G.E.M. 1957, *Intention*, Blackwell, Oxford.

Anscombe G.E.M. 1965, *The Intentionality of Sensation: A Grammatical Feature*, [w:] R.J. Butler (red.), *Analytical Philosophy* (2nd Series), Blackwell, Oxford, s. 160.

Anton G. 1899, *Ueber die Selbstwahrnehmung der Herderkrankungen des Gehirs durch den Kranken bei Rindenblindheit under Rindentaubheit*, „Archiv für Psychiatrie und Nervenkrankheitene", *32*, s. 86–127.

Arnauld A. 1641, *Fourth Set of Objections*, [w:] J. Cottingham, R. Stoofhoff, D. Murdoch, *The Philosophical Writings of Descartes*, Vol. II, 1984, Cambridge University Press, Cambridge; *Zarzuty czwarte*, [w:] R. Descartes, *Medytacje o pierwszej filozofii. Zarzuty uczonych mężów i odpowiedzi autora*, tłum.. M. i K. Ajdukiewiczowie, S. Swieżawski, Wydawnictwo Antyk, Kęty 2001.

Baars B. 1988, *A Cognitive Theory of Consciousness*, Cambridge University Press, Cambridge.

Bach-y-Rita P. 1972, *Brain Mechanisms in Sensory Substitution*, Academic Press, New York – London.

Ballard D., Feldman J. 1982, *Connectionist Models and Their Properties*, „Cognitive Science", *6*, s. 205–254.

Bechtel W., Abrahamsen A. 1991, *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*, Blackwell, Oxford.

Bennett J. 1965, *Substance, Reality and Primary Qualities*, „American Philosophical Quarterly", *2*, s. 1–17.

Bennett J. 1976, *Linguistic Behavior*, Cambridge University Press, Cambridge.

Bentham J. 1789, *Introduction to Principles of Morals and Legislation*, London; *Wprowadzenie do zasad moralności i prawodawstwa*, tłum. B. Nawroczyński, Państwowe Wydawnictwo Naukowe, Warszawa 1958.

Bick P.A., Kinsbourne M. 1987, *Auditory Hallucinations and Subvocal Speech in Schizophrenic Patients*, „American Journal of Psychiatry", *144*, s. 222–225.

Bieri P. 1990, *Commentary at the conference „The Phenomenal Mind – How Is It Possible and Why Is It Necessary?"*, Zentrum für Interdisziplinare Forschung, Bielefeld, Niemcy, Maj 14–17.

Birnbaum L., Collins G. 1984, *Opportunistic Planning and Freudian Slips*, „Proceedings, Cognitive Science Society", Boulder, CO, s. 124–127.

Bisiach E. 1988, *The (Haunted) Brain and Consciousness*, [w:] Marcel, Bisiach, 1988. Bisiach E., Luzzatti C. 1978, *Unilateral Neglect of Representational Space*, „Cortex", *14*, s. 129–133.

Bisiach E., Vallar G. 1988, *Hemineglect in Humans*, [w:] F. Boller, J. Grafman (red.), *Handbook of Neuropsychology*, Vol. 1, Elsevier, New York.

Bisiach E., Vallar G., Perani D., Papagno C., Berti A. 1986, *Unawareness of Disease Following Lesions of the Right Hemisphere: Anosognosia for Hemiplegia and Anosognosia for Hemianopia*, „Neuropsychologia", *24*, s. 471–482.

Blakemore C. 1976, *Mechanics of the Mind*, Cambridge University Press, Cambridge.

Block N. 1978, *Troubles with Funtionalism*, [w:] W. Savage (red.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minnesota Studies in the Philosophy of Science, vol. IX, s. 261–326.

Block N. 1981, *Psychologism and Behaviorism*, „Philosophical Review", *90*, s. 5–43.

Block N. 1990, *Inverted Earth*, [w:] J.E. Tomberlin (red.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, Ridgeview Publishing, Atascadero, CA, s. 53–79.

Boghossian P.A., Velleman J.D. 1989, *Colour as a Secondary Quality*, „Mind", *98*, s. 81–103.

Boghossian P.A., Velleman J.D. 1991, *Physicalist Theories of Color*, „Philosophical Review", *100*, s. 67–106.

Booth W. 1988, *Voodoo Science*, „Science", *240*, s. 274–277.

Borges J.L. 1962, *Labyrinths: Selected Stories and Other Writings*, D.A. Yates, J.E. Irby (red.), New Directions, New York; *Fikcje*, tłum. K. Piekarec i in., PIW, Warszawa 1972.

Borgia G. 1986, *Sexual Selection in Bowerbirds*, „Scientific American", *254*, s. 92–100.

Braitenberg V. 1984, *Vehicles: Experiments in Synthetic Psychology*, MIT Press – A Bradford Book, Cambridge.

Breitmeyer B.G. 1984, *Visual Masking*, Oxford University Press, Oxford.

Broad C.D. 1925, *Mind and Its Place in Nature*, Routledge & Kegan Paul, London.

Bronowski J., 1973, *Ascent of Man*, BBC Books, London; *Potęga wyobraźni*, tłum. S. Amsterdamski, Państwowy Instytut Wydawniczy, Warszawa 1988.

Brooks B.A., Yates J.T., Coleman R.D. 1980, *Perception of Images Moving at Saccadic Velocities During Saccades and During Fixation*, „Experimentol Brain Research", *40*, s. 71–78.

Byrne R., Whiten A. 1988, *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, Clarendon, Oxford.

Calvanio R., Petrone P.N., Levine D.N. 1987, *Left visual spatial neglect is both environment-centered and body-centered*, „Neurology", *37*, s. 1179–1183.

Calvin W. 1983, *The Throwing Madonna: Essays on the Brain*, McGraw-Hill, New York.

Calvin W. 1986, *The River that Flows Uphill: A Journey from the Big Bang to the Big Brain*, Sierra Club Books, San Francisco.

Calvin W. 1987, *The Brain as a Darwin Machine*, „Nature", *330*, s. 33–34.

Calvin W. 1989a, *The Cerebral Symphony: Seashore Reflections on the Structure of Consciousness*, Bantam, New York.

Calvin W. 1989b, *A Global Brain Theory*, „Science", 240, s. 1802–1803.

Campion J., Latto R., Smith Y.M. 1983, *Is Blindsight an Effect of Scattered Light, Spared Cortex, and Near-Threshold Vision?*, „Behovioral and Brain Sciences", *6*, s. 423–486.

Camus A. 1942, *Le Myth de Sisyphe*, Gallimard, Paris; *Mit Syzyfa*, tłum. J. Guze, De Agostini, Warszawa 2001.

Carruthers P. 1989, *Brute Experience*, „Journal of Philosophy", *86*, s. 258–269.

Castañeda C. 1968, *The Teachings of Don Juan: A Yaqui Way of Knowledge*, University of California Press, Berkeley; *Nauki Don Juana*, tłum. A. Szostkiewicz, Wydawnictwo Literackie, Kraków 1991.

Castaneda H.-N. 1967, *Indicators and Quasi-Indicators*, „American Philosophy Quarterly", *4*, s. 85–100.

Castaneda H.-N. 1968, *On the Logic of Attributions of Self-Knowledge to Others*, „Journal of Philosopny", *65*, s. 439–456.

Changeux J.-P., Danchin A. 1976, *Selective Stabilization of Developing Synapses as a Mechanism for the Specifications of Neuronal Networks*, „Nature", *264*, s. 705–712.

Changeux J.-P., Dehaene S. 1989, *Neuronal Models of Cognitive Functions*, „Cognition", *33*, s. 63–109.

Cheney D.L., Seyfarth R.M. 1990, *How Monkeys See the World*, University of Chicago Press, Chicago.

Cherniak C. 1986, *Minimal Rationality*, MIT Press –A Bradford Book, Cambridge, MA.

Churchland P.M. 1985, *Reduction, Qualia and the Direct Inspection of Brain States*, „Journal of Philosophy", *82*, s. 8–28.

Churchland P.M. 1990, *Knowing Qualia: A Reply to Jackson*, s. 67–76, [w:] P.M.Churchland, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press –A Bradford Book, Cambridge, MA.

Churchland P.S. 1981a, *On the Alleged Backwards Referral of Experiences and Its Relevance to the Mind-Body Problem*, „Philosophy of Science", *48*, s. 165–181.

Churchland P.S. 1981b, *The Timing of Sensations: Reply to Libet*, „Philosophy of Science", *48*, s. 492–497.

Churchland P.S. 1986, *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, MIT Press – A Bradford Book, Cambridge, MA.

Clark R.W. 1975, *The Life of Bertrand Russell*, Weidenfeld and Nicolson, London.

Cohen L.D., Kipnis D., Kunkle E.C., Kubzansky P.E. 1955, *Case Report: Observation of a Person with Congenital Insensitivity to Pain*, „Journal of Abnormal and Social Psychology",

51, s. 333–338.

Cole D. 1990, *Functionalism and Inverted Spectra*, „Synthese", *82*, s. 207–222.

Grane H., Piantanida T.P. 1983, *On Seeing Reddish Green and Yellowish Blue*, „Science", *222*, s. 1078–1060.

Crick F. 1984, *Function of the Thalamic Reticular Complex: The Searchlight Hypothesis*, „Proceedings of the National Academy of Sciences", *81*, s. 4586–4590.

Crick F., Koch C. 1990, *Towards a Neurobiological Theory of Consciousness*, „Seminars in the Neurosciences", *2*, s. 263–275.

Damasio A.R., Damasio H., Van Hoesen G.W. 1982, *Prosopagnosia: Anatomie Basis and Behavioral Mechanisms*, „Neurology", *32*, s. 331–341.

Darwin C. 1871, *The Descent of Man, and Selection in Relation to Sex*, 2 vols.,Murray, London; *O pochodzeniu człowieka*, tłum. E. Stołyhwo, S. Panek, Państwowe Wydawnictwo Rolnicze i Leśne, Warszawa 1959.

Davis W. 1985, *The Serpent and the Rainbow*, Simon & Schuster, New York.

Davis W. 1988a, *Passage of Darkness: The Ethnobiology of the Haitian Zombie*, University of North Carolina Press, Chapel Hill – London.

Davis W. 1988b, *Zombification*, „Science", *240*, s. 1715–1716.

Dawkins M.S. 1980, *Animal Suffering: The Science of Animal Welfare*, Chapman & Hall, London.

Dawkins M.S. 1987, *Minding and Mattering*, [w:] C. Blakemore, S. Greenfield (red.), *Mindwaves*, Blackwell, Oxford, s. 150–160.

Dawkins M.S. 1990, *From an Animal's Point of View: Motivation, Fitness, and Animal Welfare*, „Behavioral and Brain Sciences", *13*, s. 1–61.

Dawkins R. 1976, *The Selfish Gene*, Oxford University Press, Oxford; *Samolubny gen*, tłum. M. Skoneczny, Prószyński i S-ka, Warszawa 1996.

Dawkins R. 1982, *The Extended Phenotype*, Freeman, San Francisco; *Fenotyp rozszerzony*: *dalekosiężny gen*, tłum. J. Gliwicz, Prószyński i S-ka, Warszawa 2003.

Dawkins R. 1986, *The Blind Watchmaker*, Norton, New York; *Ślepy zegarmistrz*, tłum. A. Hoffman, Państwowy Instytut Wydawniczy, Warszawa 1994.

de Sousa R. 1976, *Rational Homunculi*, [w:] A.O. Rorty (red.), *The Identity of Persons*, University of California Press, Berkeley, s. 217–238.

Dennett D.C. 1969, *Content and Consciousness*, Routledge & Kegan Paul, London.

Dennett D.C. 1971, *Intentional Systems*, „Journal of Phiiosophy", *8*, s. 87–106.

Dennett D.C. 1974, *Why the Law of Effect Will Not Go Away*, „Journal of the Theory of Social Behaviour", *5*, s. 169–187 (przedruk [w:] Dennett, 1978a).

Dennett D.C. 1976, *Are Dreams Experiences?*, „Philosophical Review".

Dennett D.C. 1978a, *Brainstarms*, Bradford Books, Montgomery, VT.

Dennett D.C. 1978b, *Skinner Skinned*, rozdz. 4 [w:] Dennett, 1978a, s. 53–70.

Dennett D.C. 1978c, *Two Approaches to Mental Images*, rozdz. 10 [w:] Dennett, 1978a, s. 174–189.

Dennett D.C. 1978d, *Where Am I?*, rozdz. 17 [w:] Dennett, 1978a, s. 310–323.

Dennett D.C. 1979a, *On the Absence of Phenomenology*, [w:] D. Gustafson, B. Tapscott (red.), *Body, Mind and Method: Essays in Honor of Virgil Aldrich*, Reidel, Dordrecht.

Dennett D.C. 1979b, *Review of Popper and Eccles, The Self and Its Brain: An Argument for mteractionism*, „Journal of Phiiosophy", *76*, s. 91–97.

Dennett D.C. 1981a, *Reflections on „Software"*, [w:] Hofstadter, Dennett, 1981.

Dennett D.C. 1981b, *Wondering Where the Yellow Went (commentary on W. Sellars's Carus Lectures)*, „Monist", *64*, s. 102–108.

Dennett D.C. 1982a, *How to Study Human Consciousness Empirically, or Nothing Comes to Mind*, „Synthese", *59*, s. 159–180.

Dennett D.C. 1982b, *Why We Think What We Do about Why We Think What We Do: Discussion on Goodman's* '*On Thoughts without Words'*, „Cognition", *12*, s. 219–227.

Dennett D.C. 1982c, *Comments on Rorty*, „Synthese", *59*, s. 349–356.

Dennett D.C. 1982d, *Notes on Prosthetic Imagination*, „New Boston Review", June, s. 3–7.

Dennett D.C. 1983, *Intentional Systems in Cognitive Ethology: The* '*Panglossian Paradigm' Defended*, „Behavioral and Brain Sciences", *6*, s. 343–390.

Dennett D.C. 1984a, *Elbow Hoom; The Varieties of Free Will Worth Wanling*, MIT Press –A Bradford Book, Cambridge, MA.

Dennett D.C. 1984b, *Carving the Mind at Its Joints, a review of Fodor*, „The Modularity of Mind, in Contemporary Psychology", *29*, s. 285–286.

Dennett D.C. 1985a, *Can Machines Think?*, [w:] M. Shafto (red.), *How We Know*, Harper & Row, New York, s. 121–145.

Dennett D.C. 1985b, *Musie of the Hemispheres, a review of M. Gazzaniga, The Social Brain*, „New York Times Book Review", November 17, s. 53.

Dennett D.C. 1986, *Julian Jaynes' Software Archeology*, „Canadion Psychology", *27*, s. 149–154.

Dennett D.C. 1987a, *The Intentional Stonce*, MIT Press – A Bradford Book, Cambridge, MA.

Dennett D.C. 1987b, *The Logical Geography of Computational Approaches: A View from the East Pole*, [w:] M. Harnish, M. Brand (red.), *Problems in the Representation of Knowledge*, University of Arizona Press, Tucson.

Dennett D.C. 1988a, *Quining Qualia*, [w:] Marcel, Bisiach, 1988, s. 42–77.

Dennett D.C. 1988b, *When Philosophers Encounter AI*, „Daedalus", 117, s. 283–296; przedruk [w:] Graubard, 1988.

Dennett D.C. 1988c, *Out of the Armchair and Into the Field*, „Poetics Today", *9*, special issue on Interpretation in Context in Science and Culture, s. 205–222.

Dennett D.C. 1988d, *The Intentional Stance in Theory and Practice*, [w:] Whiten, Byrne, 1988, s. 180–202.

Dennett D.C. 1988e, *Science, Philosophy and Interpretation*, „Behavioral and Brain Sciences", *11*, s. 535–546.

Dennett D.C. 1988f, *Why Everyone Is a Novelist*, „Times Literary Supplement", September 16–22.

Dennett D.C. 1989a, *Why Creative Intelligence Is Hard to Find, commentary on Whiten and Byrne*, „Behovioral and Brain Sciences", *11*, s. 253.

Dennett D.C. 1989b, *The Origins of Selves*, „Cogito", *2*, s. 163–173.

Dennett D.C. 1989c, *Murmurs in the Cathedral, review of R. Penrose*, „The Emperor's New Mind, in Times Literary Supplement", September 29–October 5, s. 1066–1068.

Dennett D.C. 1989d, *Cognitive Ethology: Hunting for Bargains or a Wild Goose Chase?*, [w:] A. Montefiore, D. Noble (red.), *Goals, Own Goals and No Goals: A Debate on Goal-Directed And Intentional Behaviour*, Unwin Hyman, London.

Dennett D.C. 1990a, *Memes and the Exploitation of Imagination*, „Journal of Aesthetics and Art Criticism", *48*, s. 127–135.

Dennett D.C. 1990b, *Thinking with a Computer*, [w:] H. Barlow (red.), *Image and Understanding*, Cambridge University Press, Cambridge, s. 297–309.

Dennett D.C. 1990c, *Betting Your Life on an Algorithm, commentary on Penrose*, „Behavioral and Brain Science", *13*, s. 660.

Dennett D.C. 1990d, *The Interpretation of Texts, People, and Other Artifacts*, „Philosophy and Phenomenological Research", *50*, s. 177–194.

Dennett D.C. 1990e, *Two Black Boxes: A Fable*, Tufts University Center for Cognitive Studies Preprint, November.

Dennett D.C. 1991a, *Real Patterns*, „Journal of Philosophy", *89*, s. 27–51; *Rzeczywiste wzorce*, tłum. M. Miłkowski, [w:] M. Miłkowski, R. Poczobut (red.), *Analityczna metafizyka umysłu*, Wydawnictwo IFiS PAN, Warszawa 2008, s. 299–326.

Dennett D.C. 1991b, *Producing Future by Telling Stories*, [w:] K.M. Ford, Z. Pylyshyn (red.), *Robots Dilemma Revisited: The Frame Problem in Artificial Intelligence*, Ablex Series in Theoretical Issues in Cognitive Science, Ablex, Norwood, NJ.

Dennett D.C. 1991C, *Mother Nature versus the Walking Encyclopedia*, [w:] W. Ramsey, S. Stich, D. Rumelhart (red.), *Philosophy and Connectionist Theory*, Erlbaum, Hillsdale, NJ.

Dennett D.C. 1991d, *Two Contrasts: Folk Craft versus Folk Science and Belief versus Opinion*, [w:] J. Greenwood (red.), *The Future of Folk Psychology: Intentionality and Cognitive Science*, Cambridge University Press, Cambridge.

Dennett D.C. 1991e, *Granny's Campaign for Safe Science*, [w:] G. Rey, B. Loewer (red.), *Fodor and His Critics*, Blackwell, Oxford.

Dennett D., Kinsbourne M. 1992, *Time and the Observer: The Where and When of Consciousness in the Brain*, „Behavioral and Brain Sciences", *15*, s. 183–247.

Descartes R. 1637, *Discours de la méthode*, Paris; *Rozprawa o metodzie*, tłum. T. Żeleński-Boy, Antyk, Kęty 2002.

Descartes R. 1641, *Meditationes de prima philosophia*, Michel Soly, Paris; *Medytacje o pierwszej filozofii: zarzuty uczonych mężów i odpowiedzi autora*, tłum. M. i K. Ajdukiewiczowie, S. Swieżawski, Wydawnictwo Antyk, Kęty 2001.

Descartes R. 1662, *De homine*, Paris; *Człowiek. Opis ciała ludzkiego*, tłum. A. Bednarczyk, Państwowe Wydawnictwo Naukowe, Warszawa 1989.

Dreyfus H. 1979, *What Computers Can't Do* (2 wyd.), Harper & Row, New York.

Dreyfus H.L., Dreyfus S.E. 1988, *Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint*, [w:] Graubard, 1988.

Eccles J.C. 1985, *Mental Summation: The Timing of Voluntary Intentions by Cortical Activity*, „Behaviorol and Brain Sciences", *8*, s. 542–547.

Eco U. 1990, *After Secret Knowledge*, „Times Literary Supplement", June 22–28, s. 666, *Some Paranoid Readings*, „Times Literary Supplement", June 29–July 5, s. 694.

Edelman G. 1987, *Neural Darwinism*, Basic Books, New York.

Edelman G. 1989, *The Remembered Present: A Biological Theory of Consciousness*, Basic Books, New York.

Efron R. 1967, *The Duration of the Present*, „Proceedings of the New York Academy of Science", *8*, s. 542–543.

Eldredge N., Gould S.J. 1972, *Punctuated Equilibria: An Alternative to Phyletic Gradualism*, [w:] T.J.M. Schopf (red.), *Models in Paleobiology*, Freeman Cooper, San Francisco, s. 82–115.

Ericsson K.A., Simon H.A. 1984, *Protocol Analysis: Verbal Reports as Data*, MIT Press – A Bradford Book, Cambridge, MA.

Evans G. 1982, John McDowell (red.), *The Varieties of Reference*, Oxford University Press, Oxford.

Ewert J.-P. 1987, *The Neuroethology of Releasing Mechanisms: Prey-catching in Toads*, „Behavioral and Brain Sciences", *10*, s. 337–405.

Farah M.J. 1988, *Is Visual Imagery Really Visual? Overlooked Evidence from Neuropsychology*, „Psychological Review", *95*, s. 307–317.

Farrell B.A. 1950, *Experience*, „Mind", *59*, s. 170–198.

Fehling M., Baars B., Fisher C. 1990, *A Functional Role of Represaion in an Autonomous, Resource-constrained Agent*, [w:] *Proceedings of Twelfth Annual Conference of the Cognitive Science Society*, Erlbaum, Hillsdale, NJ.

Fehrer E., Raab D. 1962, *Reaction Time to Stimuli Masked by Metacontrast*, „Journal of Experimental Psychology", *63*, s. 143–147.

Feynman R. 1985, *Surely You're Joking, Mr. Feynman!*, Norton, New York*; „Pan raczy żartować, panie Feynman!". Przypadki ciekawego człowieka*, tłum. T. Bieroń, Znak, Kraków 2007.

Finke R.A., Pinker S., Farah M.J. 1989, *Reinterpreting Visual Patteras in Mental Imagery*, „Cognitive Science", *13*, s. 51–78.

Flanagan O. 1991. *The Science of the Mind* (2 wyd.), MIT Press – A Bradford Book, Cambridge, MA.

Flohr H. 1990, *Brain Processes and Phenomenal Consciousness: A New and Specific Hypothesis, presented at the conference „The Phenomenal Mind – How Is It Possible and Why Is It Necessary?"*, Zentrum für Interdisziplinare Forschung, Bielefeld, Niemcy, May 14–17.

Fodor J. 1975, *The Language of Thought*, Crowell, Scranton, PA.

Fodor J. 1983, *The Modularity of Mind*, MIT Press – A Bradford Book, Cambridge, MA.

Fodor J. 1990, *A Theory of Contenit, and Other Essays*, MIT Press – A Bradford Book, Cambridge, MA.

Fodor J., Pylyshyn Z. 1988, *Connectionism and Cognitive Architecture: A Critical Analysis*, „Cognilion", *28*, s. 3–71.

Fox I. 1989, *On the Nature and Cognitive Function of Phenomenal Content – Part One*, „Philosophical Topics", *17*, s. 81–117.

French R. 1991, *Subcognition and the Turing Test*, „Mind", *99*, s. 53–66.

Freud S. 1921/2012, *Ego i Id*, [w:] tegoż, *Poza zasadą przyjemności*, tłum. J. Prokopiuk, Wydawnictwo Naukowe PWN, Warszawa.

Freyd J. 1989, *Dynamie Mental Representations*, „Psychological Review", *94*, s. 427–438. Fuster J.M. 1981, *Prefrontal Cortex in Motor Control*, [w:] *Handbook of Physiology*, Section 1: *The Nervous System*, Vol. II*: Motor Control*, American Physiological Society, s. 1149–1178.

Gardner H. 1975, *The Shattered Mind*, Knopf, New York.

Gardner M. 1981, *The Laffer Curve and Other Laughs in Current Economics*, „Scientific American", *245*, December, s. 18–31; przedruk [w:] Gardner, 1986.

Gardner M. 1986, *Knotted Doughnuts and Other Mathematical Diversions*, W. H. Freeman, San Francisco.

Gazzaniga M. 1978, *Is Seeing Believing: Notes on Clinical Recovery*, [w:] S. Finger (red.), *Recovery From Brain Damage: Research and Theory*, Plenum Press, New York, s. 409–414.

Gazzaniga M. 1985, *The Social Brain: Discovering the Networks of the Mind*, Basic

Books, New York.

Gazzaniga M., Ledoux J. 1978, *The Integrated Mind*, Plenum Press, New York.

Geldard F. A. 1977, *Cutaneous Stimulis, Vitratory and Saltatory*, „Journal of Investigative Dermatology", *69*, s. 83–87.

Geldard F.A., Sherrick C.E. 1972, *The Cutaneous „Rabbit": A Perceptual Illusion*, „Science", *178*, s. 178–179.

Geldard F.A., Sherrick C.E. 1983, *The Cutaneous Saltatory Area and Its Presumed Neural Base*, „Perception and Psychophysics", *33*, s. 299–304.

Geldard F.A., Sherrick C.E. 1986, *Space, Time and Touch*, „Scientific American", *254*, s. 90–95.

Gert B. 1965, *Imagination and Verifiability*, „Philosophical Studies", *16*, s. 44–47.

Geshwind N., Fusillo M. 1966, *Color-naming Defects in Association with Alexia*, „Archives of Neurology", *15*, s. 137–146.

Gide A. 1948, *Les Faux Monnoyeurs*, Gallimard, Paris; *Fałszerze*, tłum. H. i J. Iwaszkiewiczowie, Muza, Warszawa 1948/1995.

Goodman N. 1978, *Ways of Worldmaking*, Harvester, Hassocks, Sussex; *Jak tworzymy świat*, tłum. M. Szczubiałka, Fundacja Aletheia, Warszawa 1997.

Goody J. 1977, *The Domestication of the Savage Mind*, Cambridge University Press, Cambridge.

Gould S. 1980, *The Panda's Thumb*, Norton, New York.

Gouras P. 1984, *Color Vision*, [w:] N. Osborn, J. Chader (red.), *Progress in Retinal Research*, Vol. 3, Pergamon Press, London, s. 227–261.

Graubard S.R. 1988, *The Artificial Intelligence Debate: False Starts, Heal Foundations* (a reprint of „Daedalus", *117*, Winter 1988), MIT Press, Cambridge, MA.

Grey Walter W. 1963, *Presentation to the Osler Society*, Oxford University, Oxford.

Grice H. P. 1957, *Meaning*, „Philosophical Review", *66*, s. 377–368.

Grice H. P. 1969, *Utterer's Meaning and Intentions*, „Philosophical Review", *78*, s. 147–177.

Hacking I. 1990, *Signing, review of Sacks, 1989*, „London Review of Books", April 5, s. 3–6.

Hampl P. 1989, *The Lax Habits of the Free Imagination*, „New York Times Book Review", March 5, s. 1, 37–39.

Handford M, 1987, *Where's Woldo?*, Boston, Little, Brown.

Hardin C.L. 1988, *Color for Philosophers: Unweaving the Rainbow*, Hackett, Indianapolis.

Hardin C.L. 1990, *Color and Illusion, presented at the conference „The Phenomenal Mind – How Is It Possible and Why Is It Necessary?*", Zentrum für Interdisziplinare Forschung, Bielefeld, Niemcy, May 14–17.

Harman G. 1990, *The Intrisic Quality of Experience*, [w:] J.E. Tomberlin (red.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, Ridgeview, Atascadero, CA, s. 31–52.

Harnad S. 1982, *Consciousness: An Afterthought*, „Cognition and Brain Theory", *5*, s. 29–47.

Harnad S. 1989, *Editoriai Commentary*, „Behaviorol and Brain Sciences", *12*, s. 183.

Haugeland J. 1981, *Mind Design: Philosophy, Psychology, Artificial Intelligence*, Bradford Books, Montgomery, VT.

Haugeland J. 1985, *Artificial Inlelligence: The Very Idea*, MIT Press – A Bradford Book, Cambridge, MA.

Hawking S. 1988, *A Brief History of Time*, Bantara, New York; *Krótka historia czasu*, tłum. P. Amsterdamski, Alfa, Warszawa 1990.

Hayes P. 1979, *The Naive Physics Manifesto*, [w:] D. Michie (red.), *Expert Systems in the Microelectronic Age*, Edinburgh University Press, Edinburgh; *Manifest fizyki naiwnej*, tłum. P. Pietrzak, „Przegląd Filozoficzno-Literacki", *4 (6)* / 2003, s. 49–85.

Hayes-Roth B. 1985, *A Blackboard Architecture for Control*, „Artificial Intelligence", *26*, s. 251–321.

Hebb D. 1949, *The Organization of Behavior: A Neuropsychological Theory*, Wiley, New York.

Hilbert D.R. 1987, *Color and Color Perception: A Study in Anthropocentric Realism*, Stanford University; Center for the Study of Language and Information.

Hintikka J. 1962, *Knowledge and Belief*, Cornell University Press, Ithaca.

Hinton G.E., Nowland S.J. 1987, *How Learning Can Guide Evolution, Complex Systems,*

I, Technical Report CMU-CS-86-128, Carnegie Mellon University, s. 495–502.

Hobbes T. 1651, *Leviathan*, Paris; *Lewiatan*, tłum. C. Znamierowski, Aletheia, Warszawa 2005.

Hoffman R.E. 1986, *What Can Schizophrenic „Voices" Tell Us?*, „Behavioral and Brain Sciences", s. 535–548.

Hoffman R.E., Kravitz R.E. 1987, *Feedforward Action Regulation and the Experience of Will*, „Behaviorol and Brain Sciences", *10*, s. 782–783.

Hofstadter D.R. 1981a, *The Turing Test: A Coffeehouse Conversation*, [w:] *Metamagical Themas*, „Scientific American", May 1981, przedruk [w:] Hofstadter, Dennett, 1981, s. 69–92.

Hofstadter D.R. 1981b, *Reflections* [on Nagel], [w:] Hofstadter, Dennett, 1981, s. 403–414.

Hofstadter D.R. 1983, *The Architecture of jumbo*, „Proceedings of the Second Machine Learning Workshop", Monticello, Il.

Hofstadter D.R. 1985, *On the Seeming Paradox of Mechanizing Creativity*, [w:] *Metamagical Themas*, Basic Books, New York, s. 526–546.

Hofstadter D.R., Dennett D.C. 1981, *The Mind's I: Fantasies and Reflections on Self and Soul*, Basic Books, New York, s. 191–201.

Holland J.H. 1975, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor.

Holland J.H., Holyoak K., Nisbett R.E., Thagard P.R. 1986, *Induction: Processes of Inference, Learning, and Discovery*, MIT Press – A Bradford Book, Cambridge, MA.

Honderich T. 1984, *The Time of a Conscious Sensory Experience and Mind-Brain Theories*, „Journal of Theoretical Biology", *110*, s. 115–129.

Howell R. 1979, *Fictional Objects: How They Are and How They Aren't*, [w:] D.F. Gustafson, B.L. Tapscott (red.), *Body, Mind and Method*, D. Reidel, Dordrecht, s. 241–294.

Hughlings Jackson J. 1915, *Hughlings Jackson on Aphasia and Kindred Affections of Speech*, „Brain", *38*, s. 1–190.

Hume D. 1739, *Treatise on Human Nature*, John Noon, London; *Traktat o naturze ludzkiej*, tłum. C. Znamierowski, Aletheia, Warszawa 2005.

Humphrey N. 1972, *„Interest" and „Pleasure":Two Determinanta of a Monkey's Visual Preferences*, „Perception", *1*, s. 395–416.

Humphrey N. 1976, *The Colour Currency of Nature*, [w:] T. Porter, B. Mikellides (red.), *Colour for Architecture*, Studio-Vista, London, s. 147–161, przedruk [w:] Humphrey, 1983a.

Humphrey N. 1983a, *Consciousness Regained*, Oxford University Press, Oxford.

Humphrey N. 1983b, *The Adaptiveness of Mentalism? Commentary on Dennett, 1983*, „Behavioral and Brain Sciences", *6*, s. 366.

Humphrey N. 1986, *The Inner Eye*, Faber & Faber, London.

Humphrey N. 1992, *A History of the Mind*, Simon & Schuster, New York.

Humphrey N., Dennett D.C. 1989, *Speaking for Our Selves: An Assessment of Multiple Personality Disorder*, „Raritan", *9*, s. 68–98.

Humphrey N., Keeble G. 1978, *Effects of Red Light and Loud Noise on the Rates at Which Monkeys Sample the Sensory Environment*, „Perception", *7*, s. 343.

Hundert E. 1987, *Can Neuroscience Contribute to Philosophy?*, [w:] C. Blakemore, S. Greenfield, *Mindwaves*, Blackwell, Oxford, s. 407–429 (przedruk jako rozdz. 7 Hundert, *Philosophy, Psychiatry, and Neuroscience: Three Approaches to the Mind*, Clarendon, Oxford 1989).

Huxley T. 1874, *On the Hypothesis that Animals Are Automata*, [w:] *Collected Essays*, London, 1893–1894.

Jackendoff R. 1987, *Consciousness and the Computational Mind*, MIT Press – A Bradford Book, Cambridge, MA.

Jackson F. 1982, *Epiphenomenal Qualia*, „Philosophical Quarterly", *32*, s. 127–136.

Jacob F. 1982, *The Possible and the Actual*, University of Washington Press, Seattle.

Janlert L.-E. 1985, *Studies in Knowledge Representation*, Institute of Information Processing, Umea.

Jarrell R. 1963, *The Bat-Poet*, Macmillan, New York.

Jaynes J. 1976, *The Origins of Consciousness in the Breakdown of the Bicameral Mind*, Houghton Mifflin, Boston.

Jerison H. 1973, *Evolution of the Brain and Intelligence*, Academic Press, New York.

Johnson-Laird P. 1983, *Mentol Models: Towards a Cognitive Science of Language, Inference, and Consciousness*, University Press, Cambridge.

Johnson-Laird P. 1988, *A Computational Analysis of Consciousness*, [w:] A.J. Marcel,

E. Bisiach (red.), *Consciousness in Contemporory Science*, Clarendon Press – Oxford University Press, New York – Oxford.

Julesz B. 1971, *Foundations of Cyclopean Perception*, University of Chicago Press, Chicago.

Keller H. 1908, *The World I Live In*, Century Co., New York.

Kinsbourne M. 1974, *Lateral Interactions in the Brain*, [w:] M. Kinsbourne, W.L. Smith (red.), *Hemisphere Disconnection and Cerebral Function*, Charles C. Thomas, Springfield, Il, s. 239–259.

Keller H. 1980, *Brain-based Limitations on Mind*, [w:] R.W. Rieber (red.), *Body and Mind: Past, Present and Future*, Academic Press, New York, s. 155–175.

Kinsbourne M., Hicks R.E. 1978, *Functional Cerebral Space: A Model for Overflow, Transfer and Interference Effects in Human Perfoimance: A Tutorial Review*, [w:] J. Requin (red.), *Attention and Performance*, 7, Erlbaum, Hillsdale, NJ s. 345–362.

Kinsbourne M., Warrington E.K. 1963, *Jargon Aphasia*, „Neuropsychologia", *l*, s. 27–37.

Kirman B.H., et al. 1968, *Congenital Insensitivity to Pain in an Imbecile Boy*, „Developmental Medicine and Child Neurology", *10*, s. 57–63.

Kitcher P. 1979, *Phenomenal Qualities*, „American Philosophical Quarterly", *16*, s. 123–129.

Koestler A. 1967, *The Ghost in the Machine*, Macmillan, New York.

Kohler I. 1961, *Experiments with Goggles*, „Scientific American", *206*, s. 62–86.

Kolers P.A. 1972, *Aspects of Motion Perception*, Pergamon Press, London.

Kolers P.A., von Grünau M. 1976, *Shape and Color in Apparent Motion*, „Vision Research", *16*, s. 329–335.

Kosslyn S.M. 1980, *Image and Mind*, Harvard University Press, Cambridge, MA.

Kosslyn S.M., Holtzman J.D., Gazzaniga M.S, Farah M.J. 1985, *A Computational Analysis of Mental Imagery Generation: Evidence for Functional Dissociation in Split Brain Patients*, „Journal of Experimental Psychology: General", *114*, s. 311–341.

Lackner J.R. 1988, *Some Proprioceptive Influences on the Perceptual Representation of Body Shape and Orientation*, „Brain", *111*, s. 281–297.

Langton C.G. 1989, *Artiflcial Life*, Addison–Wesley, Redwood City, CA.

Larkin S., Simon H.A. 1987, *Why a Diagram Is (Sometimes) Worth Ten Thousand Words*, „Cognitive Science", *11*, s. 65–100.

Leiber J. 1988, *„Cartesian" Linguistics?*, „Philosophia", *118*, s. 309–346.

Leiber J. 1991, *Invitation to Cognitive Science*, Blackwell, Oxford.

Leibniz G.W. 1714/1840, *La Monadologie*, [w:] J.E. Erdmann (red.), Leibniz, *Opera Philosophica*, 2 vols., Berlin; *Wyznanie wiary filozofa. Rozprawa metafizyczna. Monadologia. Zasady natury i łaski oraz inne pisma filozoficzne*, tłum. S. Cichowicz, Państwowe Wydawnictwo Naukowe, Warszawa 1969.

Levelt W. 1989, *Speaking*, MIT Press – A Bradford Book, Cambridge, MA.

Levy J., Trevarthen C. 1976, *Metacontrol of Hemispheric Function in Human Split-Brain Patients*, „Journal of Experimental Psychology: Human Perception and Performance", *3*, s. 299–311.

Lewis D. 1978, *Truth in Fiction*, „American Philosophical Quarterly", *15*, s. 37–46.

Lewis D. 1979, *Attitudes De Dieto and De Se*, „Philosophical Review", *78*, s. 513–543.

Lewis D. 1988, *What Experience Teaches, proceedings of the Russellian Society of the University of Sidney*, przedruk [w:] W. Lycan (red.), *Mind and Cognition: A Reader*, Blackwell, Oxford 1990.

Liberman A., Studdert-Kennedy M. 1977, *Phonetic Perception*, [w:] R. Held, H, Leibowitz, H.-L. Teuber (red.), *Handbook of Sensory Physiology*, Vol. 8, *Perception*, Springer-Verlag, Heidelberg.

Libet B. 1965, *Cortical Activation in Conscious and Unconscious Experience*, „Perspectives in Biology and Medicine", *9*, s. 77–86.

Libet B. 1981, *The Experimental Evidence for Subjective Referral of a Sensory Experience backwards in Time: Reply to P.S. Churchland*, „Philosophy of Science", *48*, s. 182–197.

Libet B. 1982, *Brain Stimulation in the Study of Neuronal Functions for Conscious Sensory Experiences*, „Human Neurobiology", *1*, s. 235–242.

Libet B. 1985a, *Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action*, „Behaviorol and Brain Sciences", *8*, s. 529–566.

Libet B. 1985b, *Subjective Antedating of a Sensory Experience and Mind-Brain Theories*, „Journal of Theoretical Biology", *114*, s. 563–570.

Libet B. 1987, *Are the Mental Experiences of Will and Self-control Significant for the*

Performance of a Voluntary Act?, „Behavioral and Brain Sciences", *10*, s. 783–786,

Libet B. 1989, *The Timing of a Subjective Experience*, „Behavioral and Brain Sciences", *12*, s. 183–185.

Libet B., Wright E.W., Feinstein B., Pearl D.K. 1979, *Subjective Referral of the Timing for a Conscious Sensory Experience*, „Brain", *102*, s. 193–224.

Liebmann S. 1927, *Ueber das Verhalten fahrbiger Formen bei Heligkeitsgleichtheit von Figur und Grund*, „Psychologie Forschung", *9*, s. 200–253.

Livingstone M.S., Hubel D.H. 1987, *Psychophysical Evidence for Separate Channels for the Perception of Form, Color, Movement, and Depth*, „Journal of Neuroscience", *7*, s. 346–368.

Lloyd M., Dybas H.S. 1966, *The Periodical Cicada Problem*, „Evolution", *20*, s. 132–149.

Loar B. 1990, *Phenomenal Properties*, [w:] J.E. Tomberlin (red.), *Philosophical Perspectives*, 4: *Action Theory and Philosophy of Mind*, Ridgeview, Atascadero, CA, s. 81–108.

Locke J. 1690, *Essay Concerning Human Understanding*, Basset, London; *Nowe rozważania dotyczące rozumu ludzkiego*, tłum. I. Dąmbska, Antyk Kęty 2001.

Lockwood M. 1989, *Mind, Brain and the Quantum*, Blackwell, Oxford.

Lodge D. 1988, *Nice Work*, Secker and Warburg, London; *Fajna robota*, tłum. K. Puławski, Zysk i S-ka, Poznań 1995.

Lycan W. 1973, *Inverted Spectrum*, „Ratio", *15*, s. 315–319.

Lycan W. 1990, *What Is the Subjectivity of the Mental?*, [w:] J.E. Tomberlin (red.), *Philosophical Perspectives*, 4: *Action Theory and Philosophy of Mind*, Ridgeview, Atascadero, CA, s. 109–130.

Marais E.N. 1937, *The Soul of the White Ant*, Methuen, London.

Marcel A.J. 1988, *Phenomenal Experience and Functionalism*, [w:] Marcel, Bisiach, 1988, s. 121–158.

Marcel A. 1993, *Slippage in the Unity of Consciousness*, [w:] R. Bornstein, T. Pittman (red.), *Perception Without Awareness: Cognitive, Clinical and Social Perspectives*, Guilford Press, New York.

Marcel A., Bisiach E. (red.) 1988, *Consciousness in Contemporary Science*, Oxford University Press, New York.

Margolis H. 1987, *Patterns, Thinking, and Cognition*, University of Chicago Press,

Chicago.

Margulis L. 1970, *The Origin of Eukaryotic Cells*, Yale University Press, New Haven.

Marks C. 1980, *Commissurotomy, Consciousness And Unity of Mind*, MIT Press –A Bradford Book, Cambridge, MA.

Marler P., Sherman V. 1983, *Song Structure Without Auditory Feedback: Emendations of the Auditory Template Hypothesis*, „Journal of Neuroscience", *3*, s. 517–531.

Marr D. 1982, *Vision*, Freeman, San Francisco.

Maynard Smith J. 1978, *The Evolution of Sex*, Cambridge University Press, Cambridge.

Maynard Smith J. 1989, *Sex, Games, and Evolution*, Harvester, Brighton, Sussex.

McClelland J., Rumelhart D. (red.) 1986, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, 2 vols, MIT Press – A Bradford Book, Cambridge, MA.

McCulloch W.S., Pitts W. 1943, *A Logical Calculus for the Ideas Immanent in Nervous Activity*, „Bulletin of Mathematical Biophysics", *5*, s. 115–133.

McGinn C. 1989, *Can We Solve the Mind-Body Problem?*, „Mind", *98*, s. 349–366; *Czy możemy rozwiązać problem umysł-ciało?*, tłum. M. Iwanicki i S. Judycki, [w:] M. Miłkowski, R. Poczobut (red.), *Analityczna metafizyka umysłu*, Wydawnictwo IFiS PAN, Warszawa 2008, s. 360–383.

McGinn C. 1990, *The Problem of Consciousness*, Blackwell, Oxford.

McGlynn S.M., Schacter D.L. 1989, *Unawareness of Deficits in Neuropsychological Syndromes*, „Journal of Clinical and Experimental Neuropsychology", *11*, s. 143–205.

McGurk H., Macdonald R. 1979, *Hearing Lips and Seeing Voices*, „Nature", *264*, s. 746–748.

McLuhan M. 1967, *The Medium Is the Message*, Bantam, New York.

Mellor H. 1981, *Beal Time*, Cambridge University Press, Cambridge.

Menzel E.W., Savage-Rumbaugh E.S., Lawson J. 1985, *Chimpanzee (Pan troglodytes) Spatial Problem Solving with the Use of Mirrors and Televised Equivalents of Mirrors*, „Journal of Comparalive Psychology", *99*, s. 211–217.

Millikan R. 1990, *Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox*, „Philosophical Review", *99*, s. 323–354.

Minsky M. 1975, *A Framework for Representing Knowledge*, „Memo" 3306, AI Lab,

MIT, Cambridge, MA (opubl. [w:] Haugeland, 1981. s. 95–128).

Minsky M. 1985, *The Society of Mind*, Simon & Schuster, New York.

Mishkin M., Ungerleider L.G., Macko K.A. 1983, *Object Vision and Spatial Vision: Two Cortical Pathways*, „Trends in Neuroscience, *64*, s. 370–375.

Monod J. 1972, *Chance and Necessity*, Knopf, New York; *Przypadek i konieczność: esej o filozofii biologii współczesnej*, tłum. J. Bukowski, Biblioteka „Głosu", Warszawa 1979.

Morris R.K., Rayner K., Pollatsek A. 1990, *Eye Movement Guidance in Reading: The Role of Parafoveal and Space Information*, „Journal of ExperimentaJ Psychology: Human Perception and Performance", *16*, s. 268–281.

Mountcastle V.B. 1978, *An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System*, [w:] G. Edelman, V.B. Mountcastle (red.), *The Mindful Brain*, MIT Press, Cambridge, MA, s. 7–50.

Nabokov V. 1930, *Zaschila Luzhina*, [w:] *Sovremennye Zapiski*, Paris; *Obrona Łużyna*, tłum. E. Siemaszkiewicz, Muza, Warszawa 2005.

Nagel T. 1971, *Brain Bisection and the Unity of Consciousness*, „Synthese", *22*, s. 396–413 (przedruk [w:] *Mortal Questions* [1979], Cambridge University Press, Cambridge); *Rozszczepienie mózgu a jedność świadomości*, tłum. A. Romaniuk, [w:] tegoż, *Pytania ostateczne*, Fundacja Aletheia, Warszawa 1997.

Nagel T. 1974, *What Is It Like to Be a Bat?*, „Philosophical Review", *83*, s. 435–450; *Jak to jest być nietoperzem*, tłum. A. Romaniuk, [w:] tegoż, *Pytania ostateczne*, Fundacja Aletheia, Warszawa 1997.

Nagel T. 1986, *The View from Nowhere*, Oxford Unwersity Press, Oxford; *Widok znikąd*, tłum. C. Cieśliński, Fundacja Aletheia, Warszawa 1997.

Neisser U. 1967, *Cognitive Psychology*, Appleton-Century-Crofts, New York.

Neisser U. 1981, *John Dean's Memory: A Case Study*, „Cognition", 9, s. 1–22.

Neisser U. 1988, *Five Kinds of Self-Knowledge*, „Philosophical Psychology", *1*, s. 35–39.

Nemirow L. 1990, *Physicalism and the Cognitive Role of Acquaintance*, [w:] W. Lycan (red.), *Mind and Cognition: A Reader*, Blackwell, Oxford, s. 490–499.

Neumann O. 1990, *Some Aspects of Phenomenal Consciousness and Their Possible Functional Correlates, presented at the conference „The Phenomenal Mind – How Is It Possible and Why Is It Necessary?"*, Zentrum fur Interdisziplinare Forschung, Bielefeld, Niemcy, May 14–17.

Newell A. 1973, *Production Systems: Models of Control Structures*, [w:] W.G. Chase (red.), *Visual Information Processing*, Academic Press, New York, s. 463–526.

Newell A. 1982, *The Knowledge Level*, „Artificial Intelligence", *18*, s. 81–132.

Newell A. 1988, *The Intentional Stance and the Knowledge Level*, „Behovioral and Brain Sciences", *11*, s. 520–522.

Newell A. 1990, *United Theories of Cognition*, Harvard University Press, Cambridge, MA.

Newell A., Rosenbloom P.S., ad Laird J.E. 1989, *Symbolic Architectures for Cognition*, [w:] M. Posner (red.), *Foundations of Cognitive Science*, MIT Press, Cambridge, MA. s. 93–132.

Nielsen T.I. 1963, *Volition: A New Experimental Approach*, „Scandinavian Journal of Psychology", *4*, s. 225–230.

Nilsson N, 1984, *Shakey the Computer*, „SRI Tech Report", Menlo Park, CA, SRI International.

Norman D.A., Shallice T. 1980, *Attention to Action: Willed and Automatic Control of Behavior*, Center for Human Information Processing (Technical Report No. 99), przedruk poprawiony [w:] R.J. Davidson, G.E. Schwartz, D. Shapiro (red.), 1986, *Consciousness and Self-flegulation*, Plenum Press, New York.

Norman D.A., Shallice T. 1985, *Attention to Action*, [w:] T. Shallice (red.), *Consciousness and Self-flegulation*, Plenum Press, New York.

Nottebohm F. 1984, *Birdsong as a Model in Which to Study Brain Processes Related to Learning*, „Condor", *86*, s. 227–236.

Oakley D.A. (red.) 1985, *Brain and Mind*, Methuen, London – New York.

Ornstein R., Thompson R.F. 1984, *The Amazing Brain*, Houghton Mifflin, Boston.

Pagels H. 1988, *The Dreams of Reason: The Computer and the Rise of the Sciences of Complexity*, Simon & Schuster, New York.

Papert S. 1988, *One AI or Many?*, „Daedalus", Winter, s. 1–14.

Parfit D. 1984, *Reasons and Persons*, Clarendon Press, Oxford.

Pears D. 1984, *Motivated Irrationality*, Clarendon Press, Oxford.

Penfield W. 1958, *The Excitable Cortex in Conscious Man*, Liverpool University Press, Liverpool.

Penrose R. 1989, *The Emperor's New Mind*, Oxford University Press, Oxford; *Nowy umysł cesarza*, tłum. P. Amsterdamski, Wydawnictwo Naukowe PWN, Warszawa 2000.

Perlis 1991, *Intentionality and Defaults*, [w:] K.M. Ford, P.J. Hayes (red.), *Reasoning Agents in a Dynamic World*, JAI Press, Greenwich, CT.

Perry J. 1979, *The Problem of the Essential Indexical*, „Nous", *13*, s. 3–21.

Pinker S., Bloom P. 1990, *Natural Language and Natural Selection*, „Behavioral and Brain Sciences", *13*, s. 707–784.

Pollatsek A., Rayner K., Collins W.E. 1984, *Integrating Pictorial Information Across Eye Movements*, „Journal of Experimental Psychology: General", *113*, s. 426–442.

Pöppel E. 1985, *Grenzen des Bewusstseins*, Deutsche Verlags-Anstal, Stuttgart.

Pöppel E. 1988 (tłum. Pöppel. 1985), *Mindworks: Time and Conscious Experience*, Harcourt Brace Jovanovich, New York.

Popper K.R., Eccles J.C. 1977, *The Self and Its Brain*, Springer-Verlag, Berlin; *Mózg i jaźń*, tłum. P. Jaśkowski, t. 1–3, Protext, Poznań 1999.

Powers L. 1978, *Knowledge by Deduction*, „Philosophical Review", *87*, s. 337–371.

Putnam H. 1965, *Brains and Behavior*, [w:] R.J. Butler (red.), *Analitical Philosophy*, Second Series, Blackwell, Oxford, s. 1–19.

Putnam H. 1988, *Much Ado About Not Very Much*, „Daedalus", *117*, Winter, przedruk [w:] Graubard, 1988.

Pylyshyn Z. 1979, *Do Mental Events Have Durations?*, „Behaviorol and Brain Sciences", *2*, s. 277–278.

Quine W.V.O. 1969, *Natural Kinds*, [w:] *Ontological Relativity and Other Essays*, Columbia University Press, New York, s. 114–138.

Ramachandran V.S. 1985, *Guest Editorial*, „Perception", *14*, s. 97–103.

Ramachandran V.S. 1991, *2-B or not 2-B: That Is the Question*, [w:] R.L. Gregory, J. Harris, P. Heard, D. Rose, C. Cronly-Dillon (red.), *The Artful Brain*, Oxford University Press, Oxford.

Ramachandran V.S., Gregory R.L., 1991, *Perceptual Filling in of Artificially Induced Scotomas in Human Vision*, „Nature", *350 (6320)*, s. 699–702.

Raphael B. 1976, *The Thinking Computer: Mind Inside Matter*, Freeman, San Francisco.

Reddy D.R., Erman L.D., Fennel R.D., Neeły R.B. 1973, *The HEARSAY-II Speech Understanding System: An Example of the Recognition Process*, „Proceedings of the International Joint Conference on Artificial Intelligence", Stanford, s. 185–194.

Reingold E.M., Merikle P.M. 1990, *On the Interrelatedness of Theory and Measurement in the Study of Unconscious Processes*, „Mind and Language", *5*, s. 9–28.

Reisberg D., Chambers D., 1991, *Neither Pictures nor Propositions: What Can We Learn from a Mental Image?*, „Canadian Journal of Psychology", *45*, s. 336–352.

Richards R.J. 1987, *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*, University of Chicago Press, Chicago.

Ristau C. 1991, *Cognitive Ethology: The Minds of Other Animals: Essays in Honor of Donald R. Griffin*, Erlbaum, Hillsdale, NJ.

Rizzolati G., Gentilucci M., Matelli M. 1985, *Selective Spatial Attention: One Center, One Circuit, or Many Circuits?*, [w:] M.I. Posner, O.S.M.Marin (red.), *Attention and Performance XI*, Erlbaum, Hillsdale, NJ.

Rorty R. 1970, *Incorrigibility as the Mark of the Mental*, „Journal of Philosophy", *67*, s. 399–424.

Rorty R. 1982a, *Contemporary Philosophy of Mind*, „Synthese", *53*, s. 323–348.

Rorty R. 1982b, *Comments on Dennett*, „Synthese", *53*, s. 181–187.

Rosenbloom P.S., Laird J.E., Newell A. 1987, *Knowledge-Level Learning in Soar*, „Proceedings of AAAl", Morgan Kaufman, Los Altos, CA.

Rosenthal D. 1986, *Two Concepts of Consciousness*, „Philosophical Studies", *49*, s. 329–359.

Rosenthal D. 1989, *Thinking That One Thinks*, „ZIF Report" No. 11, Research Group on Mind and Brain, Perspectives in Theoretical Psychology and the Philosophy of Mind, Zentrum für Interdisziplinare Forschung, Bielefeld, Niemcy.

Rosenthal D. 1990a, *Why Are Verbally Expressed Thoughts Conscious?*, „ZIF Report" No. 32, Zentrum für Interdisziplinare Forschung, Bielefeld, Niemcy.

Rosenthal D. 1990b, *A Theory of Consciousness*, „ZIF Report" No. 40, Zentrum für Interdisziplinare Forschung, Bielefeld, Niemcy.

Rozin P. 1976, *The Evolution of Intelligence and Access to the Cognitive Unconscious*, „Progress in Psychobiology and Physiological Psychology", *6*, s. 245–280,

Rozin P. 1982, *Human Food Selection: The Interation of Biology, Culture and Individual*

Experience, [w:] L.M. Barker (red.), *The Psychobiology of Human Food Selection*, Avi Publishing Co, Westport, CT.

Rozin P., Fallon A.E. 1987, *A Perspective on Disgust*, „Psychological Review", *94*, s. 23–47.

Russell B. 1927, *The Analysis of Matter*, Allen and Unwin, London.

Ryle G. 1949, *The Concept of Mind*, Hutchinson, London; *Czym jest umysł*, tłum. W. Marciszewski, Państwowe Wydawnictwo Naukowe, Warszawa 1970.

Ryle G. 1979, *On Thinking*, K. Kolenda (red.), *Rowman and Littlefield*, Totowa, NJ.

Sacks O. 1985, *The Man Who Mistook His Wife or His Hat*, Summit Books, New York; *Mężczyzna, który pomylił swoją żonę z kapeluszem*, tłum. B. Lindenberg, Zysk i S-ka, Poznań 2001.

Sacks O. 1989, *Seeing Voices*, University of California Press, Berkeley; *Zobaczyć głos: podróż do świata ciszy*, tłum. A. Małaczyński, Zysk i S-ka 1998..

Sandeval E. 1991, *Towards a Logic of Dynamie Frames*, [w:] K.M. Ford, J. Hayes (red.), *Reasoning Agents in a Dynamie World*, JAI Press, Greenwich, CT.

Sanford D. 1975, *Infinity and Vagueness*, „Philosophical Review", *84*, s. 520–535.

Sartre J.-P. 1943, *L'Etre et le Néant*, Gallimard, Paris; *Byt i nicość*, tłum. J. Kiełbasa i in., Zielona Sowa, Kraków 2007.

Schank R. 1991, *Tell Me a Story*, Scribners, New York.

Schank R., Abelson R. 1977, *Scripts, Plans. Goals and Understanding: An Inquiry into Human Knowledge Structures*, Erlbaum, Hillsdale, NJ.

Schull J. 1990, *Are Species Intelligent?*, „Behavioral and Brain Sciences", *13*, s. 63–108.

Searle J. 1980, *Minds, Brains, and Programs*, „Behovioral and Brain Sciences", *3*, s. 417–458; *Umysły, mózgi i programy*, tłum. B. Chwedeńczuk, [w:] B. Chwedeńczuk (red.), *Filozofia umysłu. Fragementy filozofii analitycznej*, Fundacja Aletheia – Wydawnictwo Spacja, Warszawa 1995.

Searle J. 1982, *The Myth of the Computer: An Exchange*, „New York Review of Books", June 24, s. 56–57.

Searle J. 1983, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge.

Searle J. 1984, *Panel Discussion: Has Artificial Intelligence Research Illuminated Human*

Thinking?, [w:] H. Pagels (red.), *Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer*, Annals of the New York Academy of Sciences, 426.

Searle J. 1988a, *Turing the Chinese Room*, [w:] T. Singh (red.), *Synthesis of Science and Religion, Critical Essays and Dialogues*, Bhaktivedenta Institute, San Francisco.

Searle J. 1988b, *The Realistic Stance*, „Behavioral and Brain Sciences", *11*, s. 527–529.

Searle J. 1990a, *Consciousness, Explanatory Inversion, and Cognitive Science*, „Behavioral and Brain Sciences", *13*, s. 585–642.

Searle J. 1990b, *Is the Brain's Mind a Computer Program?*, „Scientific American", *262*, s. 26–31.

Selfridge O. 1959, *Pandemonium: A Paradigm for Learning*, Symposium on the Mechonizotion of Thought Processes, HM Stationery Office, London.

Searle J., Rękopis, Tracking and Trailing.

Sellars W. 1963, *Empiricism and the Philosophy of Mind*, [w:] *Science, Perception and Reality*, Routledge & Kegan Paul, London; *Empiryzm i filozofia umysłu*, tłum. J. Gryz, [w:] B. Stanosz (red.), *Empiryzm współczesny*, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 1991, s. 173–257.

Sellars W. 1981, *Foundations for a Metaphysics of Pure Process*, (the Carus Lectures), „Monist", *64*, s. 3–90.

Shallice T. 1972, *Dual Functions of Consciousness*, „Psychological Review", *79*, s. 383–393.

Shallice T. 1978, *The Dominant Action System: An Information-Processing Approach to Consciousness*, [w:] K.S. Pope, J.L. Singer (red.), *The Stream of Consciousness*, Plenum, New York, s. 148–164.

Shallice T. 1988, *From Neuropsychology to Mental Structure*, Cambridge University Press, Cambridge.

Sharpe T. 1977, *The Great Pursuit*, Secker and Warburg, London.

Shepard R.N. 1964, *Circularity in Judgments of Relative Pitch*, „Journal of the Acoustical Society of America", *36*, s. 2346–2353.

Shepard R.N., Cooper L.A. 1982, *Mental Images and Their Transformations*, MIT Press – A Bradford Book, Cambridge, MA.

Shepard R.N., Metzler J. 1971, *Mental Rotation of Three-Dimensional Objects*, „Science", *171*, s. 701–703.

Shoemaker S. 1969, *Time Without Change*, „Journal of Philosophy", *66*, s. 363–381.

Shoemaker S. 1975, *Functionalism and Qualia*, „Synthese", *27*, s. 291–315.

Shoemaker S. 1981, *Absent Qualia are Impossible – A Reply to Block*, „Philosophical Review", *90*, s. 581–599.

Shoemaker S. 1988, *Qualia and Consciousness*, Tufts University Philosophy Department Colloquium.

Siegel R.K., West L.J. (red.) 1975, *Hallucinations: Behavior, Experience and Theory*, Wiley, New York.

Simon H.A., Kaplan C.A. 1989, *Foundations of Cognitive Science*, [w:] Posner (red.), *Foundations of Cognitive Science*, MIT Press, Cambridge, MA.

Smolensky P. 1988, *On the Proper Treatment of Connectionism*, „Behavioral and Brain Sciences", *11*, s. 1–74.

Smullyan R.M. 1981, *An Epistemological Nightmare*, [w:] Hofstadter, Dennett, 1981, s. 415–427, przedruk [w:] tegoż, *Philosophical Fantasies*, St. Martin's Press, New York 1982.

Smythies J.R. 1954, *Analysis of Projection*, „British Journal of Philosophy of Science", *5*, s. 120–133.

Snyder D.M. 1988, *On the Time of a Conscious Peripheral Sensation*, „Journal of Theoretical Biology", *130*, s. 253–254.

Sperber D., Wilson D. 1986, *Relevance: A Theory of Communication*, Harvard University Press, Cambridge, MA; *Relewancja: komunikacja i poznanie*, tłum. zbiorowe, Tertium, Kraków 2011.

Sperling G. 1960, *The Information Available in Brief Visual Presentations*, „Psychological Monographs", *74*, No. 11.

Sperry R.W. 1977, *Forebrain Commissurotomy and Conscious Awareness*, „The Journal of Medicine and Philosophy", *2*, s. 101–126.

Spillman L., Werner J.S. 1990, *Visual Perception: The Neurophysiological Foundations*, Academic Press, San Diego.

Spinoza B. 1677, *Tractatus de Intellectus Emendatione*; *Traktat o uzdrowieniu rozumu...* [w:] tegoż; *Pisma wczesne*, tłum. L. Kołakowski, Wydawnictwo Naukowe PWN, Warszawa 2009.

Stafford S.P. 1983, *On The Origin of the Intentional Stance*, Tufts University Working

Paper in Cognitive Science, CCM 83–1.

Stalnaker R. 1984, *Inquiry*, MIT Press – A Bradford Book, Cambridge, MA.

Stix G. 1991, *Reach Out*, „Scientific American", *264*, s. 134.

Stoerig P., Cowey A. 1990, *Wavelength Sensitivity in Blindsight*, „Nature", *342*, s. 916–918.

Stoll C. 1989, *The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage*, Doubleday, New York; *Kukułcze jajo*, tłum. T. Hornowski, Dom Wydawniczy Rebis, Poznań 1998.

Straight H.S. 1976, *Comprehension versus Production in Linguistic Theory*, „Foundations of Language", *14*, s. 525–540.

Stratton G.M. 1896, *Some Preliminary Experiments on Vision Without Inversion of the Retinal Image*, „Psychology Review", *3*, s. 611–617.

Strawson G. 1989, *Red and „Red"*, „Synthese", *78*, s. 193–232.

Strawson P.F. 1962, *Freedom and Resentment*, „Proceedings of the British Academy", przedruk [w:] P.F. Strawson (red.), *Studies in the Philosophy of Thought and Action*, Oxford University Press, Oxford 1968.

Taylor D.M. 1966, *The Incommunicability of Content*, „Mind", *75*, s. 527–541.

Thompson D'Arcy W. 1917, *On Growth and Form*, Cambridge University Press, Cambridge.

Thompson E., Palacios A., Varela F. 1992, *Ways of Coloring. Comparative Color Vision as a Case Study for Cognitive Science*, „Behavioral and Brain Sciences", *15 (01)*, s. 1–26.

Tranel D., Damasio A.R. 1988, *Non-conscious Face Recognition in Patients with Face Agnosia*, „Behavioral Brain Research", *30*, s. 235–249.

Tranel D., Damasio A.R., Damasio H. 1988, *Intact Recognition of Facial Expression, Gender, and Age in Patients with Impaired Recognition of Face Identity*, „Neurology", *38*, s. 690–696.

Treisman A. 1988, *Features and Objects: The Fourteenth Bartlett Memorial Lecture*, „Quarterly Journal of Experimental Psychology", *40A*, s. 201–237.

Treisman A., Gelade G. 1980, *A Feature-integration Theory of Attention*, „Cognitive Psychology", *12*, s. 97–136.

Treisman A., Sato S. 1990, *Conjunction Search Revisited*, „Journal of Experimental

Psychology: Human Perception and Performance", *16*, s. 459–478.

Treisman A., Souther J. 1985, *Search Asymmetry: A Diagnostic for Preattentive Processing of Separable Features*, „Journal of Experimental Psychology: General", *114*, s. 285–310.

Turing A. 1950, *Computing Machinery and Intelligence*, „Mind", *59*, s. 433–460; *Maszyny liczące a inteligencja*, tłum. B. Chwedeńczuk, [w:] B. Chwedeńczuk (red.), *Filozofia umysłu*, Aletheia, Warszawa 1995, s. 271-300.

Tye M. 1986, *The Subjective Qualities of Experience*, „Mind", *95*, s. 1–17.

Uttal W.R. 1979, *Do Central Nonlinearities Exist?*, „Behavioral and Brain Sciences", *2*, s. 286.

Van der Waals H.G., Roelofs C.O. 1930, *Optische Scheinbewegung*, „Zeitschrift für Psychologie und Physiologie des Sinnesorgane", *114*, s. 241–288; *115* (1931), s. 91–190.

Van Essen D.C. 1979, *Visual Areas of the Mammalian Cerebral Cortex*, „Annual Review of Neuroscience", *2*, s. 227–263.

Van Gulick R. 1988, *Consciousness, Intrinsic Intentionality, and Self-understanding Machines*, [w:] Marcel, Bisiach, 1988, s. 78–100.

Van Gulick R. 1989, *What Difference Does Consciousness Make?*, „Phliosophical Topics", *17*, s. 211–230.

Van Gulick R. 1990, *Understanding the Phenomenal Mind: Are We All Just Armadillos?, presented at the conference „The Phenomenal Mind – How Is It Possible and Why Is It Necessary?*, Zentrum für Interdisziplinare Forschung, Bielefeld, Niemcy, May 14–17.

Van Tuijl H.F.J.M. 1975, *A New Visual Illusion: Neonlike Color Spreading and Complementary Color Induction between Subjective Contours*, „Acta Psychologica", *39*, s. 441–445.

Vendler Z. 1972, *Res Cogitans*, Cornell University Press, Ithaca.

Vendler Z. 1984, *The Matter of Minds*, Clarendon Press, Oxford.

Von der Malsburg C. 1985, *Nervous Structures with Dynamical Links*, „Berichte der Bunsen-Gesellschaftr Physikalische Chemie", *89*, s. 703–710.

Von Uexküll J. 1909, *Umwelt und Innenwelt der Tiere*, Berlin – Jena.

Vosberg R., Fraser N., Guehl J. 1960, *Imagery Sequence in Sensory Deprivation*, „Archives of General Psychiatry", *2*, s. 356–357.

Walton K. 1973, *Pictures and Make Believe*, „Philosophical Review", *82*, s. 283–319.

Walton K. 1978, *Fearing Fiction*, „Journal of Philosophy", *75*, s. 6–27.

Warren R.M. 1970, *Perceptual Restoration of Missing Speech Sounds*, „Science", *167*, s. 392–393.

Wasserman G.S. 1985, *Neural/Mental Chronometry and Chronotheology*, „Behavioral and Brain Sciences, B", s. 556–557.

Weiskrantz L. 1986, *Blindsight: A Case Study and Implications*, Oxford University Press, Oxford.

Weiskrantz L. 1988, *Some Contributions of Neuropsychology of Vision and Memory to the Problem of Consciousness*, [w:] Marcel, Bisiach, 1988, s. 183–199.

Weiskrantz L. 1989, *Panel discussion on consciousness*, „European Brain and Behavior Society", September 1989, Turin.

Weiskrantz L. 1990, *Outlooks for Blindsight: Explicit Methodologies for Implicit Processes* (The Ferrier Lecture), „Proceedings of the Royal Society London, B", *239*, s. 247–278.

Welch R.B. 1978, *Perceptual Modification: Adapting (o Altered Sensory Environments*, Academic Press, New York.

Wertheimer M. 1912, *Experimentelle Studienuber das Sehen von Bewegung*, „Zeitschrift für Psychologie", *61*, s. 161–265.

White S.L. 1986, *The Curse of the Qualia*, „Synthese", *68*, s. 333–368.

Whiten A., Byrne R. 1988, *Toward the Next Generation in Data Quality: A New Survey of Primate Tactical Deception*, „Behavioral and Brain Sciences", *11*, s. 267–273.

Wiener N. 1948, *Cybernetics: or Control and Communication in the Animal and the Machine*, Technology Press, Cambridge; *Cybernetyka, czyli sterowanie i komunikacja w zwierzęciu i maszynie*, tłum. J. Mościcki, Państwowe Wydawnictwo Naukowe, Warszawa 1971.

Wilkes K.V. 1988, *Real People*, Oxford University Press, Oxford.

Wilsson L. 1974, *Observations and Experiments on the Ethology of the European Beaver*, Viltrevy, „Swedish Wildlife", *8*, s. 115–266.

Winograd T. 1972, *Understanding Natural Language*, Academic Press, New York.

Wittgenstein L. 1953, *Philosophical Investigations*, Blackwell Oxford; *Dociekania filozoficzne*, tłum. B. Wolniewicz, Wydawnictwo Naukowe PWN, Warszawa 2000.

Wolfe J.M. 1990, *Three Aspects of the Parallel Guidance of Visual Attention*, „Proceedings of the Cognitive Science Society", Erlbaum, Hillsdale, NJ, s. 1048–1049.

Yonas A. 1981, *Infants Responses to Optical Information for Collision*, [w:] R.N. Aslin, J.R. Alberts, M.R. Peterson (red.), „Development of Perception: Psychofaiological Perspectives", Vol. 2: *The Visual System*, Academic Press, New York.

Young J.Z. 1965a, *The Organization of a Memory System*, „Proceedings Royal Society London [Biology]", *163*, s. 285–320.

Young J.Z. 1965b, *A Model of the Brain*, Clarendon, Oxford; *Model mózgu*, tłum. S. Bogusławski, Państwowe Wydawnictwo Naukowe, Warszawa 1968.

Young J.Z. 1979, *Learning as a Process of Selection*, „Journal of the Royal Society of Medicine", *72*, s. 801–804.

Zajonc R., Markus H. 1984, *Affect and Cognition: The Hard Interface*, [w:] C. Izard, J. Kagan, R. Zajonc (red.), *Emotion, Cognition and Behavior*, Cambridge University Press, Cambridge, s. 73–102.

Zeki S.M., Shipp S. 1988, *The Functional Logic of Cortical Connections*, „Nature", *335*, s. 311–317.

Zihl J. 1980, *„Blindsight": Improvement of Visually Guided Eye Movements by Systematic Practice in Patients with Cerebral Blindness*, „Neuropsychologica", *18*, s. 71–77.

Zihl J. 1981, *Recovery of Visual Functions in Patients with Cerebral Blindness*, „Experimental Brain Research", *44*, s. 159–169.

Bibliografia do Posłowia

Akins K. 1996. *Lost the Plot? Reconstructing Dennett's Multiple Drafts*, „Mind and Language", *7 (1)*, s. 1–43.

Baars B.J. 1988. *A Cognitive Theory of Consciousness*, Cambridge University Press, Cambridge – New York.

Baars B.J. 1997. *In the Theater of Consciousness. The Workspace of the Mind*, Oxford University Press, New York – Oxford.

Bauby J.-D. 2008. *Skafander i motyl*, tłum. K. Rutkowski, Wyd. 2, Słowo/Obraz Terytoria.

Boly M., Seth A.K., Wilke M., Ingmundson P., Baars B.J., Laureys S., Edelman D., Tsuchiya N. 2013. *Consciousness in Humans and Non-Human Animals: Recent Advances and*

Future Directions, „Frontiers in Psychology", *4*. doi:10.3389/fpsyg.2013.00625.

Chalmers D.J. 2010. *Świadomy umysł: w poszukiwaniu teorii fundamentalnej*, tłum. M. Miłkowski, Wydawnictwo Naukowe PWN, Warszawa.

Dennett D.C. 1988. *Quining Qualia*, [w:] A. Marcel, E. Bisiach (red.), *Consciousness in Modern Science*, Oxford University Press, Oxford.

Dennett D.C. 2003. *Naprawdę przekonani: strategia intencjonalna i dlaczego ona działa*, tłum. M. Miłkowski, „Przegląd Filozoficzno-Literacki", *4 (6)*, s. 87–109.

Dennett D.C. 2007. *Słodkie sny: filozoficzne przeszkody na drodze do nauki o świadomości*, tłum. M. Miłkowski, Prószyński i S-ka, Warszawa.

Dennett D.C. 2008. *Rzeczywiste wzorce*, [w:] M. Miłkowski, R. Poczobut (red.), *Analityczna metafizyka umysłu*, tłum. M. Miłkowski, Wydawnictwo IFiS PAN, Warszawa, s. 299–326.

Dennett D.C. 2015. *Dźwignie wyobraźni i inne narzędzia do myślenia*, tłum. Ł. Kurek, Copernicus Center Press, Kraków.

Gallagher S., Zahavi D. 2015. *Fenomenologiczny umysł*, tłum. M. Pokropski, Wydawnictwo Naukowe PWN, Warszawa.

Górska U., Koculak M., Brocka M., Binder M. 2014. *Zaburzenia świadomości – perspektywa kliniczna i etyczna*, „Aktualnosci Neurologiczne", *14 (3)*, s. 190–198. doi:10.15557/AN.2014.0022.

Hurlburt R.T., Schwitzgebel E. 2007. *Describing Inner Experience? Proponent Meets Skeptic*, MIT Press, Cambridge, MA.

Jackendoff R.S. 1990. *Consciousness and the Computational Mind*, MIT Press, Cambridge, MA.

Johnson-Laird Ph.N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*, Harvard University Press, Cambridge, MA.

Koch Ch. 2008. *Neurobiologia na tropie świadomości*, tłum. G. Hess, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa.

Metzinger Th. 2000. *Neural Correlates of Consciousness*, MIT Press, Cambridge, MA.

Metzinger Th. 2003. *Being No One. The Self-Model Theory of Subjectivity*, MIT Press, Cambridge, MA.

Nishimoto S., Vu A.T., Naselaris Th., Benjamini Y., Bin Y., Gallant J.L. 2011. *Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies*, „Current

Biology : CB", *21 (19)*, s. 1641–1646. doi:10.1016/j.cub.2011.08.031.

Paprzycka K. 2005. *O możliwości antyredukcjonizmu*, Semper, Warszawa.

Poczobut R. 2009. *Między redukcją a emergencją: spór o miejsce umysłu w świecie fizycznym*, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław.

Prinz J.J. 2012. *The Conscious Brain: How Attention Engenders Experience*n, Oxford University Press, Oxford – New York.

Revonsuo A. 2006. *Inner Presence: Consciousness as a Biological Phenomenon*, MIT Press, Cambridge, MA.

Rosenthal D. 2005. *Consciousness and Mind*, Oxford University Press, Oxford – New York.

Seth A.K., Dienes Z., Cleeremans A., Overgaard M., Pessoa L. 2008. *Measuring Consciousness: Relating Behavioural and Neurophysiological Approaches*, „Trends in Cognitive Sciences", *12 (8)*, s. 314–321. doi:10.1016/j.tics.2008.04.008.

Tononi G. 2004. *An Information Integration Theory of Consciousness*, „BMC Neuroscience", *5 (1)*. doi:10.1186/1471-2202-5-42.

Tyszka T., (red.) 1995. *Czy powrót do introspekcji? Zbieranie i analiza danych słownych*, Wydawnictwo Naukowe PWN, Warszawa.

Wierzchoń M. 2013. *Granice świadomości: w poszukiwaniu poznawczego modelu subiektywności*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.

Autor dziękuje za udostępnienie ilustracji:

Ryc. 2.3: © 1969 Harvey Comics Entertainment, Inc.

Ryc. 2.4: © 1975 Sidney Harris – „American Scientist".

Ryc. 4.1: Z *Shakey the Computer*, Nils Nilsson, Copyright International. Reprinted with permission. 1984. SRI

Ryc. 3: Z *The ThinkingComputer: Mind Inside Matter*, Bertram Raphael. Copyright © 1976 by W. H. Freeman and Company. Reprinted with permission.

Ryc. 5.1: Z *A Brief History of Time*. Copyright © 1988 by Stephen W. Hawking. Bantam, Doubleday, Dell Publishing Group, Inc. Reprinted with permission.

Ryc. 5.5 i 5.6: Z *Mathematical Games*, Martin Gardner, „Scientific American", December 1981. Reprinted with permission.

Ryc. 5.7: Z *Foundations of Cyclopean Vision*, Bela Julesz. Copyright © 1971 by University of Chicago Press. Reprinted with permission.

Ryc. 7.4: Oparto na Macro Computer, an interactive computer simulation of a computer, developed by Steve Barney at the Curricular Software Studio, Tufts University.

Ryc. 8.1: Z SpeaJdng: From Intention to Articulation. Willem J.M. Levelt. Copyright © 1989 by Massachusetts Institute of Technology. Reprinted with permission.

Ryc. 9.1: Z *The Architecture of Cognition*, John R. Andersen, Cambridge, Mass.: Harvard University Press, Copyright © 1983 by the President and Fellows of Harvard College. Reprinted with permission.

Ryc. 10.1: Z *Brainstorms*, Daniel Dennett. Copyright © 1978 by Bradford Books, Publishers. Published by MIT Press. Reprinted with permission.

Ryc. 10.7: Rysunek Gahana Wilsona © 1990 by The New Yorker Magazine, Inc. Reprinted with permission.

Ryc. 11.3: Z *Brain Mechanisms in Sensory Substitution*, Paul Bach-y-Rita. Copyright © 1972 Academic Press, Inc. Reprinted with permission.

Ryc. 11.4: Z „Nature", vol. 221, s. 963–964. Copyright © 1969 Macmillan Magazines Ltd. Reprinted with permission.

Przypisy

[1] Pojęcie *eksplozja kombinatoryczna* pochodzi z informatyki, ale fenomen ten został rozpoznany na długo przed komputerami, na przykład w bajce o władcy, od którego wieśniak, uratowawszy mu życie, zażądał ziarenek ryżu: jednego na pierwszym polu szachowym, dwóch na drugim, czterech na trzecim i tak dalej, podwajając liczbę ziarenek na każdym z 64 pól. Okazało się, że władca jest winien podstępnemu wieśniakowi miliony miliardów ziarenek ryżu (a dokładnie 2^{64}-1). Jeszcze lepszym przykładem jest problem, z jakim zetknęli się francuscy „aleatoryczni" powieściopisarze, którzy po pierwszym rozdziale wprowadzali zasadę, że czytelnik rzuca monetą, po czym czyta odpowiednio rozdział 2a lub 2b, następnie czyta odpowiednio rozdział 3aa, 3ab, 3ba lub 3bb i tak dalej, za każdym razem rzucając monetą. Owi powieściopisarze szybko zrozumieli, że lepiej zminimalizować ilość momentów rzutu monetą, niż poddać się eksplozji fikcji, która uniemożliwiłaby komukolwiek przeniesienie całej „książki" z księgarni do domu.

[2] Rozwój systemów „rzeczywistości wirtualnej" dla celów rozrywkowych i badawczych przeżywa obecnie swój boom. Stan tej sztuki jest imponujący: elektroniczne rękawiczki oraz przekonujący interfejs do „manipulowania" wirtualnymi obiektami czy zakładane na głowę projektory wizualne, które pozwalają eksplorować złożone wirtualne rzeczywistości. Ograniczenia tych systemów są jednak oczywiste i ukazują sedno sprawy: silne iluzje mogą być podtrzymane tylko poprzez kombinacje kopii fizycznych oraz schematyzację (które zapewniają stosunkowo ubogą reprezentację wirtualną). A nawet w najlepszym wypadku są doświadczeniami surrealistycznymi, a nie czymś, co można by choć na moment pomylić z rzeczywistością. Jeśli naprawdę chcesz oszukać kogoś tak, aby myślał, że jest w klatce z gorylem, zapewnienie sobie pomocy aktora w stroju goryla będzie najlepszym wyjściem na długie lata.

[3] Zainteresowanych bardziej szczegółowym omówieniem kwestii związanych z wolną wolą, kontrolą, czytaniem w myślach oraz przewidywaniem odsyłam do mojej książki *Elbow Room: The Varieties of Free Will Worth Wanting* (1984), szczególnie do rozdziałów 3 i 4.

[4] Praktyka pokazuje, że istnieje większe prawdopodobieństwo stworzenia ciekawej historii, jeśli będziemy częściej przychylać się do odpowiedzi twierdzących, stawiając linię podziału alfabetu pomiędzy literami *p* i *q*.

[5] Zainteresowanych dalszą dyskusją na ten temat odsyłam do rozdziału 4 w mojej książce *Elbow Room...* (1984).

[6] Ang. *animate* – ożywiony, *inanimate* – nieożywiony (przyp. tłum.).

[7] Ang. *The buck stops here* (przyp. tłum.).

[8] Kilka dzielnych dusz (które niewątpliwie nie mogą przeciwstawić się takiemu kategoryzowaniu!) opiera się temu trendowi: *The Ghost in the Machine* (1967) Arthura Koestlera oraz *The Self and Its Brain* (1977) Karla Poppera i Johna Ecclesa to książki niepodważalnie eminentnych autorów, a dwa kolejne ikonoklastyczne i dziwnie wnikliwe obrony dualizmu to *Res Cogitans* (1972) i *The Matter of Minds* (1984) Zeno Vendlera.

[9] Odsyłam do mojej recenzji tej książki: *Murmurs in the Cathedral* (Dennett 1989c).

[10] Eccles twierdził, że niefizyczny umysł składa się z milionów „psychonów", które współdziałają z milionami „dendronów" (układami piramidowymi) w korze mózgowej; każdy psychon odpowiada z grubsza temu, co Kartezjusz czy Hume nazwaliby „ideami" – na przykład idea koloru czerwonego, idea okrągłości, idea ciepła – ale poza tą szczątkową analizą Eccles nie

ma nic do powiedzenia o częściach, czynnościach, zasadach działania czy innych cechach niefizycznego umysłu.

[11] Fascynująco status teorii Landa relacjonuje filozof C.L. Hardin w aneksie do swojej książki *Color for Philosophers: Unweaving the Rainbow* (1988).

[12] Prawdę mówiąc, Kartezjusz również miał taki pogląd na zwierzęta. Uważał, że zwierzęta to po prostu wyszukane maszyny. Ludzkie ciała, a nawet mózgi, także były dla niego jedynie maszynami. Tylko niemechaniczne, niefizyczne umysły były tym, co czyniło istoty ludzkie (i tylko istoty ludzkie) inteligentnymi i świadomymi. Był to tak naprawdę łagodny punkt widzenia, który zostałby w dużej mierze obroniony przez dzisiejszych zoologów, ale dla współczesnych Kartezjuszowi był on zbyt radykalny, więc karykaturowali go na wszystkie wyobrażalne sposoby i po prostu z niego drwili. Wieki później owo szkalowanie nadal jest radośnie propagowane przez tych, dla których prospekt mechanicznego wytłumaczenia świadomości jest nie do pojęcia – a przynajmniej nie do przyjęcia. Zainteresowanych osobliwą relacją na ten temat odsyłam do Leibera (1988).

[13] Tego terminu używa się w ten sposób w anglojęzycznej tradycji filozoficznej; w Polsce i w innych krajach Europy kontynentalnej „fenomenologia" to raczej teoria lub koncepcja zjawisk przeżywanych, a nie same te zjawiska. Dla uproszczenia jednak w niniejszym przekładzie stosujemy ten termin w użyciu typowym dla angielszczyzny (przyp. red. nauk.).

[14] Obrońcy fenomenologii zwykle podkreślają jej odrębność wobec zwykłych metod introspekcyjnych; wprowadzenie do współczesnych metod fenomenologicznych w filozofii umysłu i kognitywistyce przedstawiają Sh. Gallagher i D. Zahavi w książce *Fenomenologiczny umysł*, przeł. M. Pokropski, Wydawnictwo Naukowe PWN, Warszawa 2015 (przyp. red. nauk.).

[15] Owo retoryczne pytanie zakłada według niektórych gromką odpowiedź: „Nic!". Na przykład McGinn (1989/2008) podpiera swoją kapitulancką odpowiedź badaniem dostępnych wariantów, ignorując możliwości, które rozwinę w kolejnych rozdziałach.

[16] Dlaczego dźwięk A poniżej środkowego C i dźwięk A powyżej środkowego C (czyli o oktawę wyższy) brzmią *podobnie*? Co sprawia, że oba są dźwiękami A? Jaka niewyrażalna cecha wysokości dźwięku je łączy? Kiedy dwa tony są oddalone od siebie o oktawę (i wówczas brzmią dla nas „tak samo, ale inaczej"), podstawowa częstotliwość jednego z nich jest dokładnie podwójną podstawową częstotliwością drugiego. Standardowe A poniżej środkowego C wibruje 220 razy na sekundę; A o oktawę wyżej („A koncertowe") wibruje 440 razy na sekundę. Zagrane jednocześnie, nuty oddalone od siebie o jedną bądź więcej oktaw będą zgodne. Czy wyjaśnia to zagadkę niewyrażalnego pokrewieństwa? Absolutnie nie. Dlaczego dźwięki zgodne w ten sposób miałyby w *ten* sposób podobnie brzmieć? Cóż, dźwięki, które nie są ze sobą zgodne, nie brzmią podobnie w *ten* sposób, ale mogą być podobne w innym sensie (na przykład mieć podobną barwę), co można wyjaśnić poprzez związek między częstotliwością a powstającymi wibracjami. Opisawszy różne podobieństwa dźwięków oraz porównawszy ich właściwości fizyczne, a także ich wpływ na nasz system słuchowy, możemy dość precyzyjnie przewidzieć, jak będą dla nas brzmieć nowe dźwięki (wytwarzane na przykład przez elektroniczne syntezatory). Jeśli wszystko to nie wyjaśnia niewyrażalnego pokrewieństwa, cóż pozostaje do wyjaśnienia? (Tym popularnym tematem zajmiemy się dość szczegółowo w rozdziale 12).

[17] Ang. *I see* (przyp. tłum.).

[18] Klasyczne rozwinięcie tego tematu, razem z późniejszym jego ugruntowaniem, choć różnej jakości, przynoszą Wittgensteina *Dociekania filozoficzne* (1953/2000).

[19] Pewien neurochirurg opowiadał mi kiedyś, jak operował mózg młodego mężczyzny z epilepsją. Pacjent był całkowicie świadom, jak zawsze w przypadku tego rodzaju operacji poddany tylko miejscowemu znieczuleniu, a chirurg delikatnie eksplorował jego korę mózgową,

upewniając się, że części, które wstępnie postanowiono usunąć, nie są niezbędne, stymulując je elektrycznie i pytając pacjenta, co czuje. Niektóre ze stymulacji powodowały wizualne błyśnięcia lub podnoszenie ręki, inne pewnego rodzaju poczucie brzęczenia, ale jedno z miejsc spowodowało uszczęśliwioną wypowiedź pacjenta: „To *Out Ta Get Me* Guns N'Roses, mojego ulubionego zespołu heavymetalowego!".

Zapytałem neurochirurga, czy poprosił pacjenta o zaśpiewanie lub zanucenie piosenki tak, jak ją słyszał, gdyż byłoby fascynujące dowiedzieć się, jak „wierne" było wywołane wspomnienie. Czy byłoby dokładnie w tej samej tonacji i tempie jak na nagraniu? Taki utwór (w przeciwieństwie do *Cichej nocy*) ma jedną kanoniczną wersję, więc moglibyśmy nałożyć nagranie nucenia pacjenta na standardowe nagranie i porównać je ze sobą. Niestety, mimo że operacja była nagrywana na taśmę magnetofonową, chirurg nie poprosił o to pacjenta. „Dlaczego nie?" – zapytałem, a on na to: „Nie znoszę muzyki rockowej!".

Później neurochirurg wspomniał, że znów będzie operował tego samego pacjenta, a ja wyraziłem nadzieję, że spróbuje ponownie pobudzić muzykę rockową w jego głowie i jednak poprosi mężczyznę o jej zaśpiewanie. „Nie mogę tego zrobić – powiedział chirurg – bo właśnie tę część będę wycinał". „To część, w której znajduje się ognisko epileptyczne?" – zapytałem, a on odpowiedział: „Nie, już ci mówiłem – nie znoszę muzyki rockowej!".

Opisana technika chirurgiczna została wprowadzona przez Wildera Penfielda wiele lat temu i obrazowo przedstawiona w *The Excitable Cortex in Conscious Man* (Penfield 1958).

[20] Literatura dotycząca ewolucyjnego objaśnienia bólu pełna jest niesamowicie krótkowzrocznych argumentów. Pewien autor uważa, że nie może istnieć żadne ewolucyjne wyjaśnienie bólu, ponieważ niektóre obezwładniające jego rodzaje, jak na przykład ten związany z kamieniami żółciowymi, wywołują alarm, z którym nikt nie był w stanie nic zrobić aż do czasów współczesnej medycyny. Żaden jaskiniowiec nie osiągnął reprodukcyjnego zysku z ataku pęcherzyka żółciowego, więc ból – a przynajmniej niektóre jego rodzaje – jest ewolucyjną tajemnicą. Autor ten jednak ignoruje prosty fakt, że aby mieć odpowiedni system bólowy, który będzie mógł ostrzec przed takimi kryzysami możliwymi do uniknięcia, jak szpon lub kieł wbity w brzuch, trzeba zapewne dostać premię – którą doceni się dużo później – w postaci systemu ostrzegającego o sytuacjach kryzysowych pozostających poza kontrolą. Analogicznie, jest wiele wewnętrznych stanów, o których chcielibyśmy się dziś dowiadywać za pomocą ostrzeżenia bólowego (na przykład stany przedrakowe), ale o których nie wiemy prawdopodobnie dlatego, że nasza ewolucyjna przeszłość nie obejmowała zwiększonej możliwości przetrwania związanej z wymaganymi obwodami nerwowymi (gdyby miały się one pojawić poprzez mutację).

[21] „Co pomyślałby Marsjanin, gdyby zobaczył śmiejącego się człowieka? Musi to wyglądać strasznie: widok rozjuszonych gestów, odrzucanych kończyn i tułowia falującego pod wpływem szalonych wygibasów" (Minsky 1985, s. 280).

[22] W ostatniej sztuce Moliera, klasycznej komedii *Chory z urojenia* (1673), Argan, tytułowy hipochondryk, rozwiązuje swoje problemy, „stając się" lekarzem, aby sam móc siebie leczyć. Nie są do tego wymagane żadne studia – jedynie odrobina poprzekręcanej łaciny. Podczas parodii egzaminu ustnego zostaje mu zadane pytanie: „Dlaczego opium usypia ludzi?". Kandydat na lekarza odpowiada: „Ponieważ ma *virtus dormitiva* – czyli »moc powodującą sen«". „*Bene, bene, bene, bene respondere*", mówi chór. Świetna odpowiedź! Ależ to pouczające! Cóż za spostrzeżenie! Natomiast w duchu bardziej współczesnym moglibyśmy zapytać: „Co takiego w Cheryl Tiegs sprawia, że jest tak fotogeniczna?".

[23] Więcej o czerwonej i zielonej plamie można się dowiedzieć z Crane i Piantanida (1983) oraz Hardin (1988); o znikającej granicy kolorów, czyli o efekcie Liebmanna (1927) – ze Spillmann i Werner (1990); o dźwięku Sheparda pisze Shepard (1964); o efekcie Pinokia –

Lackner (1988). Więcej o prozopagnozji można przeczytać w: Damasio, Damasio i Van Hoesen (1982), Tranel i Damasio (1988), Tranel, Damasio i Damasio (1988).

[24] Ta oraz następna część rozdziału opierają się na kilku moich wcześniejszych opisach metodologicznych podstaw heterofenomenologii: Dennett (1978c, 1982a).

[25] Kilka lat temu Wade Davis, młody antropolog z Harvardu, oświadczył, że rozszyfrował zagadkę zombi wudu i w swojej książce *The Serpent and the Rainbow* (1985) opisał napój neurofarmakologiczny przygotowywany przez praktyków wudu, który przypuszczalnie wprowadza człowieka w stan podobny do śmierci; po kilku dniach bycia zakopanym żywcem ci biedacy bywają czasem odkopywani i karmieni halucynogenem, wywołującym dezorientację i amnezję. W rezultacie halucynogenu lub uszkodzenia mózgu spowodowanego niedotlenieniem pod ziemią mogą rzeczywiście szurać nogami w sposób przypominający zombi z filmów, a czasem są również brani do niewoli. Z powodu sensacyjnej natury twierdzeń Davisa (oraz filmu luźno opartego na książce) jego odkrycia spotkały się z nutą sceptycyzmu w pewnych kręgach, ale zarzuty te zostały odparte w jego drugiej, bardziej naukowej publikacji *Passage of Darkness: The Ethnobiology of the Haitian Zombie* (1988). Zobacz również: Booth (1988) i Davis (1988b).

[26] W rozdziale *Jak zmienić zdanie* z książki *Brainstorms* (1978a) przyjmuję konwencjonalne użycie terminu „opinia", co pozwala mi rozróżnić właściwe przekonania od innych stanów związanych raczej z językiem, które nazywam „opiniami". Zwierzęta, które nie mają języka, mogą mieć przekonania, ale nie opinie. Ludzie mają obie te rzeczy, lecz jeśli uważasz, że jutro jest piątek, to według mojej terminologii powinna to być twoja opinia, że jutro jest piątek. Nie jest to rodzaj stanu poznawczego, który moglibyśmy mieć bez języka. Nie będę zakładał obycia czytelnika z tą różnicą, ale zamierzam rościć sobie prawo do odwoływania się do obu tych kategorii.

[27] Przypomina to trudności, z jakimi zmagają się fizycy badający *osobliwość*, czyli punkt, w którym właśnie z powodu jego niewymiarowości niektóre wielkości są nieskończone (co wynika z ich definicji). Wiąże się to z istnieniem czarnych dziur, ale wpływa również na interpretację bardziej przyziemnych zagadnień. Roger Penrose rozważa to, jak zastosować równania Lorentza i Maxwella do cząstek. „Zgodnie z równaniem Lorentza, rozważając ruch cząstek, musimy zbadać pole elektromagnetyczne dokładnie w punkcie, w którym znajduje się naładowana cząstka (w ten sposób otrzymujemy siłę działającą na tę cząstkę). Jaki punkt należy wziąć, jeśli cząstka ma niezerowy promień? Czy należy znaleźć pole w »środku« cząstki, czy też może należy obliczyć średnią po całej jej »powierzchni«? [...] Być może zatem lepiej jest przyjąć model cząstek punktowych. To jednak też prowadzi do trudności, bowiem w tym wypadku pole w bezpośrednim otoczeniu cząstki dąży do nieskończoności" (Penrose 1989/2000, s. 216).

[28] Byłoby szaleństwem zaprzeczać temu, że głowa jest centralą, jednak byli i tacy. W roku 1800 Philippe Pinel zrelacjonował dziwny przypadek mężczyzny, który popadł w „prawdziwe delirium spowodowane horrorem rewolucji. Utrata zdrowego rozsądku wyróżnia się u niego czymś niezwykłym: uważa, że ścięto mu głowę, która została pośpiesznie rzucona na stos głów innych ofiar, a następnie sędziowie, zbyt późno żałując swego okrutnego czynu, rozkazali, aby pozbierano głowy i ponownie je przytwierdzono do odpowiednich ciał. Jednak w wyniku jakiegoś błędu przytwierdzono mu głowę innego nieszczęśnika. Ów wymysł, że podmieniono mu głowę, zajmuje go dzień i noc. [...] – Popatrz na moje zęby! – powtarzał bez przerwy – kiedyś były wspaniałe, a te są zepsute! *Moje* usta były zdrowe, a *te* są zarażone! Cóż za różnica pomiędzy tymi włosami a tamtymi, które miałem, zanim zamieniono mi głowę!". *Traité médico-philosophique sur l'aliénation mentale, ou La manie*, Chez Richard, Caile et Ravier, Paris 1800, s. 66–67. (Dziękuję Dorze Weiner za zwrócenie mojej uwagi na ten fascynujący przypadek).

[29] Reaganomika – polityka finansowa amerykańskiego prezydenta Ronalda Reagana charakteryzująca się obniżaniem podatków i zmniejszaniem wydatków budżetowych (przyp. tłum.).

[30] Ponieważ trzeba byłoby użyć dosyć starego stereoskopu, chyba lepiej dziś posłużyć się tekturowymi okularami do rzeczywistości wirtualnej. Produkuje się takie do smartfonów (wzór do samodzielnego wykonania jest na stronie https://developers.google.com/cardboard/ (przyp. red. nauk.).

[31] Jeszcze bardziej uderzający przykład to eksperyment, w którym osoba badana jest oszukana za pomocą luster i myśli, że obserwuje swoją własną rękę rysującą linię, gdy tak naprawdę obserwuje rękę współpracownika eksperymentatora. W tym przypadku „oczy wygrywają" aż tak, że proces redakcyjny w mózgu jest oszukany na tyle, iż stwierdza, że ręka osoby badanej jest poruszana na siłę; osoba twierdzi, że czuje „nacisk" uniemożliwiający „jej" ręce poruszanie się tam, gdzie powinna (Nielsen 1963).

[32] Operacjonizm to (mniej więcej) strategia wyrażona w następujący sposób: „Jeśli nie potrafisz znaleźć różnicy, to różnica nie istnieje", lub, jak można często usłyszeć: „Jeśli chodzi jak kaczka i kwacze jak kaczka, to musi być kaczka". Analizę silnych i słabych punktów operacjonizmu przeprowadza Dennett (1985a).

[33] W korze mózgowej jest obszar zwany „okolicą środkowo-skroniową", który reaguje na ruch (oraz pozorny ruch). Załóżmy więc, że pewna aktywność w tym obszarze to dochodzenie do wniosku, że nastąpił ruch pomiędzy punktami. W modelu wielokrotnych szkiców nie pojawia się pytanie, czy jest to wniosek przed przeżyciem czy po nim. Innymi słowy, błędem byłoby zapytać, czy aktywność w okolicy środkowo-skroniowej była „reakcją na świadome przeżycie" (według historyka orwellowskiego) czy „decyzją o reprezentacji ruchu" (według stalinowskiego redaktora).

[34] Hobbes był czujny na problemy związane z tą kwestią: „Gdyby bowiem barwy i dźwięki były w ciałach czy przedmiotach, które je wywołują, to nie można by było ich oddzielić od tych ciał, jak to się dzieje, gdy patrzymy przez szkła, czy też, gdy się odbije echo; wiemy wówczas, że rzecz, którą dostrzegamy, jest w jednym miejscu, obraz zaś w innym" (Hobbes 1651/1954, s. 9). Ten fragment poddawany jest jednak kilku innym interpretacjom.

[35] Smythies 1954. Ten heroiczny tekst pokazuje, jak trudno było myśleć o tych kwestiach wcale nie tak dawno temu. Autor dokładnie udowadnia fałszywość podręcznikowej wersji teorii rzutowania, a w podsumowaniu cytuje Bertranda Russella, który obala ten sam pomysł: „Ten, kto akceptuje przyczynową teorię percepcji, jest zmuszony stwierdzić, że przedmioty postrzeżenia są w naszych głowach, ponieważ pojawiają się na końcu przyczynowego łańcucha zdarzeń, prowadzącego, przestrzennie, od przedmiotu do mózgu podmiotu postrzegającego. Nie możemy przypuszczać, że na końcu tego procesu ostatni ze skutków nagle przeskakuje z powrotem do początku niczym rozciągnięta lina, która pęka" (Russell 1927).

[36] „To tak, jakby nasz feenomanista zmieniony w feenomenologa miał w całym zamieszaniu chwycić się fortelu wymyślenia przestrzeni dla boga, czyli nieba, aby jego ukochany Feenoman mógł w nim przebywać, przestrzeni wystarczająco *rzeczywistej* dla wyznawcy, a jednocześnie na tyle odległej i tajemniczej, aby ukryć w niej Feenomana przed sceptykami. Przestrzeń fenomenalna to raj obrazów umysłowych, jednak jeśli obrazy umysłowe okazują się *prawdziwe*, mogą wygodnie przebywać w fizycznej przestrzeni w naszych mózgach, a jeśli okażą się nieprawdziwe, mogą przebywać, razem ze świętym Mikołajem, w logicznej przestrzeni fikcji" (Dennett 1978a, s. 186).

[37] Filozof Jay Rosenberg zwrócił moją uwagę na to, że Kant dostrzega w tym mądrość,

twierdząc, iż w doświadczeniu to, co *für mich* („dla mnie”), oraz to, co *an sich* („samo w sobie”), są tym samym.

[38] Filozof Ned Block opowiedział mi kiedyś swoje przeżycie jako osoby badanej w teście „lateralizacji”. Patrzył przed siebie w konkretny punkt, a co jakiś czas słowo (bądź coś niebędącego słowem, na przykład GHRPE) pojawiało się po lewej lub prawej stronie punktu, na który patrzył. Jego zadaniem było wciśnięcie przycisku, gdy bodziec był słowem. Jego czasy reakcji były dłuższe dla słów pojawiających się po lewej stronie (i docierających najpierw do prawej półkuli), potwierdzając hipotezę, że miał, jak większość ludzi, silną lateralizację funkcji językowych w lewej półkuli. Nie było to zaskoczeniem dla Blocka; zainteresowała go natomiast „fenomenologia: słowa pojawiające się z lewej strony wydawały się trochę zamazane”. Zapytałem go, czy uważał, że słowa były trudniejsze do zidentyfikowania, ponieważ wydawały się zamazane, czy też wydawały się zamazane, gdyż były trudniejsze do zidentyfikowania. Przyznał, że nie jest w stanie rozróżnić tych „przeciwstawnych” przyczynowych opisów dokonywania osądu.

[39] Ten sposób myślenia o owej kwestii przyszedł mi do głowy po przeczytaniu Snydera (1988), chociaż jego sposób spojrzenia na te problemy jest trochę inny niż mój.

[40] Argumenty i analizy w tym rozdziale (i pewne ich elementy w rozdziale poprzednim) są rozwinięciem materiału, którym zajęli się Dennett i Kinsbourne (1992).

[41] Nie oznacza to, że mózg nigdy nie korzysta ze „wspomnień buforowych”, aby złagodzić różnice wynikające z wewnętrznych procesów zachodzących w mózgu i asynchronicznego świata zewnętrznego. Oczywistym przykładem jest „pamięć echoiczna”, dzięki której na krótką chwilę przechowujemy schematy bodźców, podczas gdy mózg rozpoczyna ich przetwarzanie (Sperling 1960; Neisser 1967; również Newell, Rosenbloom i Laird 1989, s. 107).

[42] Spieszę z wyjaśnieniem, że podkoloryzowałem tu historię. Francis Rawdon-Hastings, pierwszy markiz Hastings i drugi hrabia Moira, był głównym dowódcą w Bengalu i gubernatorem generalnym Indii w 1815 roku, jednak nie mam bladego pojęcia, jak i kiedy został poinformowany o bitwie pod Nowym Orleanem.

[43] Tego rodzaju „stempel” może być z założenia dodany do nośnika treści na którymkolwiek z etapów podróży; jeśli wszystkie materiały docierające do danego miejsca pochodzą z tej samej lokalizacji oraz przybyły tą samą trasą, z taką samą prędkością, ich „czas wyjazdu” może być później do nich dołączony poprzez zwykłe odjęcie pewnej stałej od czasu przybycia do celu. Jest to inżynieryjna możliwość prawdopodobnie wykorzystywana przez nasz mózg w celu pewnego rodzaju automatycznego dopasowania do standardowych czasów podróży.

[44] Jak zauważa Uttal (1979), to rozróżnienie *jest* szeroko uznawane przez neuronaukowców: „Istota wielu badań w dziedzinie kodowania sensorycznego może zostać przedstawiona w postaci jednej ważnej idei – dowolny potencjalny kod może reprezentować dowolny wymiar percepcyjny; nie ma potrzeby izomorficznej relacji między danymi neurologicznymi i psychofizycznymi. Przestrzeń może reprezentować czas, czas może reprezentować przestrzeń, miejsce może reprezentować jakość i oczywiście nieliniowe funkcje neuronowe mogą równie dobrze reprezentować liniowe bądź nieliniowe funkcje psychofizyczne” (s. 286). Jest to idea powszechnie znana, jednak wkrótce zobaczymy, że niektórzy teoretycy rozumieją ją na opak; sposób, w jaki doszukują się w niej sensu, polega na milczącym, ponownym wprowadzeniu niepotrzebnego „izomorfizmu” do niejasno wyobrażanego procesu tłumaczenia czy „rzutowania”, który miałby następować w świadomości.

[45] Zob. także Pylyshyn 1979 (s. 178): „Nikt [...] nie chce *dosłownie* mówić o takich własnościach fizycznych zdarzeń umysłowych jak kolor, wielkość, masa itp. [...], chociaż

opisujemy je jako coś, co *reprezentuje* takie własności (lub ma treść przeżywaną). Na przykład nie powiedzielibyśmy o myśli (lub obrazie), że była duża lub czerwona, a jedynie, że była myślą o czymś dużym i czerwonym (lub że był to obraz czegoś dużego i czerwonego). [...] Jest więc ciekawe, że z taką pewnością mówimy o *trwaniu* zdarzenia umysłowego".

[46] Jak pisze psycholog Robert Efron (1976, s. 721): „Gdy po raz pierwszy obserwujemy przedmiot w środku pola widzenia, nie pojawia się przelotne przeżycie tegoż obiektu na obrzeżu pola widzenia, a dopiero potem coraz bardziej w jego środku. [...] Analogicznie, gdy przenosimy uwagę z jednego przedmiotu świadomości na drugi, nie pojawia się przeżycie »rosnącej« dokładności nowego przedmiotu świadomości – po prostu postrzegamy nowy przedmiot".

[47] Ang. Strategic Defense Initiative (SDI) – plan budowy systemu obronnego przeciwko potencjalnym atakom nuklearnym ze strony ZSRR rozwijany przez Stany Zjednoczone w latach 1983–1991 (przyp. tłum.).

[48] Zobacz również, jak Libet odrzuca bardziej umiarkowaną sugestię MacKaya (1981, s. 195; 1985b, s. 568). Z drugiej strony ostateczne podsumowanie Libeta w 1981 roku jest niekonkluzywne: „Według mnie [...] czasowa rozbieżność sprawia duży problem teorii identyczności, ale nie jest on nie do pokonania" (s. 196). Prawdopodobnie byłby zdecydowanie nie do pokonania w interpretacji *rzutowania* wstecz, gdyż zakłada ona prekognicję, przyczynowość wsteczną lub coś równie niedorzecznego i niespotykanego. Co więcej, później Libet opisuje owe trudności nie do pokonania w sposób, który wymaga bodaj łagodniejszego podejścia: „Mimo że hipoteza opóźniania i datowania wstecz nie odłącza rzeczywistego czasu przeżycia od czasu wytworzenia go w układzie nerwowym, to eliminuje potrzebę równoczesności *subiektywnego umiejscowienia w czasie* danego doświadczenia i rzeczywistego, obiektywnego czasu przeżycia" (Libet 1985b, s. 569). Być może entuzjastyczne wsparcie sir Johna Ecclesa dla radykalnej, dualistycznej interpretacji tych rezultatów odwróciło uwagę Libeta (i jego krytyków) od łagodnej tezy, której od czasu do czasu jest obrońcą.

[49] We wcześniejszym artykule Libet uznał możliwość procesu orwellowskiego i założył, że może istnieć znaczna różnica między nieświadomymi zdarzeniami umysłowymi a świadomymi, lecz efemerycznymi zdarzeniami umysłowymi: „Może równie dobrze istnieć bezpośredni, ale efemeryczny rodzaj przeżycia świadomego, który nie zostaje zachowany w pamięci na świadomych poziomach przeżycia. Jeśli jednak takie przeżycia istnieją, ich treść miałaby bezpośrednie znaczenie tylko w późniejszych, nieświadomych procesach umysłowych, chociaż, jak inne doświadczenia nieświadome, mogą one grać pośrednią rolę w późniejszych przeżyciach świadomych" (Libet 1965, s. 78).

[50] Harnard dostrzega nierozwiązywalny problem pomiaru, jednak zaprzecza temu, co twierdzę ja – że faktycznie nie ma takiego momentu: „Introspekcja może powiedzieć nam jedynie, kiedy zdarzenie *zdawało się* następować lub które ze zdarzeń *wydawało się* zdarzać pierwsze. Nie ma niezależnego sposobu potwierdzenia, że realne umiejscowienie w czasie było w rzeczywistości tym, co się wydawało. Niewspółmierność jest problemem metodologicznym, a nie metafizycznym" (Harnard 1989, s. 183).

[51] Można wyciągnąć wniosek, że uważam, iż wszystko, z czego korzystam w opowiadaniu mojej historii, jest według mnie słuszne – lub idzie w słusznym kierunku – jednak na podstawie faktu, że nie zawarłem w mojej opowieści pewnych teorii czy szczegółów teorii, nie wolno wnioskować, że uważam je za błędne. Nie można też sądzić, że tylko dlatego, iż korzystam z kilku detali pewnej teorii, uważam, że cała jej reszta jest uzasadniona. Odnosi się to również do wykorzystanych tu moich wcześniejszych prac na ten temat.

[52] John Maynard Smith jest czołowym teoretykiem tego zagadnienia i poza jego klasycznym dziełem *The Evolution of Sex* (1978) istnieje kilka świetnych artykułów dotyczących

problemów pojęciowych w jego zbiorze esejów *Sex, Games and Evolution* (1989). Krótki przegląd tej kwestii również w: Dawkins (1976/1996, s. 71–73).

[53] Koncepcja wielofunkcyjnych neuronów nie jest nowa, ale ostatnio zyskuje zwolenników: „Niedwuznaczne są mniej więcej równoczesne konkatenacje wyjść czy sygnałów neuronalnych, a nie wyjścia indywidualnych neuronów. Zbieżność różnych konkatenacji dwuznacznych sygnałów na każdym kolejnym poziomie częściowo rozwiązałaby tę dwuznaczność, tak jak zbieżność dwuznacznych definicji wyznacza unikatowe lub niemal unikatowe rozwiązania w krzyżówce" (Dennett 1969, s. 56). „[...] nie ma unikatowej struktury czy kombinacji grup odpowiadających danej kategorii czy wzorcowi wyjścia. Zamiast tego więcej niż jedna kombinacja grup neuronalnych może wygenerować określone wyjście, a dana pojedyncza grupa może brać udział w więcej niż jednym rodzaju funkcji sygnalizacyjnej. Ta właściwość grup neuronalnych w repertuarach, zwana *degeneracją*, stanowi podstawę generalizujących zdolności map wielobieżnych" (Edelman 1989, s. 50). Już Hebb w pionierskiej pracy *The Organization of Behavior: A Neuropsychological Theory* (1949) podkreślał wagę tej cechy architektury, w której każdy węzeł przyczynia się do wielu różnych treści. Cecha ta leży u podstaw „równoległego przetwarzania rozproszonego" czy „koneksjonizmu". Jednak „wielofunkcyjność" znaczy coś więcej; na ogólniejszym poziomie analizy otrzymamy pełne systemy o wyspecjalizowanych rolach, ale też mogące być składnikami konstrukcji bardziej uniwersalnych.

[54] Chyba jako pierwszy na analogię pomiędzy żachwami a profesorami wskazał neuronaukowiec Rodolfo Llinás.

[55] Ten element konstrukcji przypomina niecałkowicie niezawodny sposób odróżniania pudełek od piramid stosowany przez Shakeya. Shakey nie jest więc zupełnie niebiologiczny; biosfera posiada wiele takich gadżetów. Jest jednak prawdą, że system „wzrokowy" Shakeya w ogóle nie jest dobrym modelem widzenia u jakiegokolwiek gatunku. Nie to było jego celem.

[56] Podstawowe ujęcie tej kwestii jest już u Darwina i wczesnych propagatorów jego teorii (Richards 1987). Neuroanatom John Z. Young (1965a, 1965b/1968) utorował drogę selekcjonistycznej teorii pamięci (zob. też Young 1979). Filozoficzną wersję podstawowej argumentacji wraz ze szkicem szczegółów przedstawiłem w mojej pracy doktorskiej na Oxfordzie w 1965 roku. Jej okrojona wersja stała się trzecim rozdziałem książki *Content and Consciousness* (1969). John Holland (1975) i inni zajmujący się sztuczną inteligencją opracowali „algorytmy genetyczne" dla samoorganizujących lub samouczących się systemów (zob. też Holland, Holyoak, Nisbett i Thagard 1986), a Jean-Pierre Changeux (Changeux i Danchin 1976; Changeux i Dehaene 1989) sporządził dosyć szczegółowy model nerwowy. Neurobiolog William Calvin (1987, 1989a) w inny (i przystępniejszy) sposób naświetla te kwestie w swojej teorii ewolucji w mózgu. Zobacz również jego jasną i wnikliwą recenzję (Calvin 1989b) z pracy *Neural Darwinism* Geralda Edelmana (1987). Niedawno Edelman opublikował też książkę *The Remembered Present: A Biological Theory of Consciousness* (1989).

[57] Ang. *fight, flee, feed or mate*.

[58] Jest to problem *treści umysłowej* lub *intencjonalności*, będący podstawowym problem w filozofii umysłu, a proponowane jego rozwiązania są zwykle kontrowersyjne. Moje rozwiązanie znajduje się w książce *The International Stance* (1987a).

[59] Kilku odważnych teoretyków twierdzi inaczej. Na przykład Jerry Fodor (1975) uważa, że wszystkie pojęcia, jakie człowiek kiedykolwiek może mieć, muszą być dane w momencie narodzin, a następnie wyzwolone czy odkryte przez konkretne epizody „uczenia się". Tak więc Arystoteles miał pojęcie samolotu w swoim mózgu, jak również pojęcie roweru – tylko nigdy nie miał okazji ich użyć! Osobom wybuchającym śmiechem na tak absurdalny

pomysł Fodor odpowiada, że immunolodzy śmiali się kiedyś z pomysłu, że ludzie – na przykład Arystoteles – rodzą się z milionami różnych przeciwciał, łącznie z przeciwciałami na określone związki, które pojawiły się w naturze dopiero w XX wieku, a teraz już się nie śmieją; okazało się to prawdą. Problem z tym pomysłem, w jego zastosowaniu zarówno w immunologii, jak i psychologii, jest taki, że jego radykalne wersje są w sposób oczywisty nieprawdziwe, a wersje umiarkowane są nie do odróżnienia od przeciwnych poglądów. W systemie immunologicznym istnieje kombinatoryczna reakcja – nie *każda* odpowiedź immunologiczna to reakcja jednoznaczna między pojedynczymi typami istniejących wcześniej przeciwciał; analogicznie, być może Arystoteles miał wrodzone pojęcie *samolotu*, jednak czy również miał pojęcie *szerokokadłubowego jumbo jeta*? A pojęcie *przelotu z Bostonu do Londynu ze zniżką wynikającą z zakupienia biletu z dużym wyprzedzeniem*? Do czasu, aż poznamy odpowiedzi na te pytania w obu dziedzinach, okazuje się, że w jednej i drugiej istnieje coś w rodzaju uczenia się oraz coś w rodzaju wrodzonych pojęć.

[60] Jest to oczywiście aluzja do teorii Paula Grice'a, zwanej teorią nienaturalnego znaczenia (Grice 1957, 1969); jednak nową teorię komunikacji, która zastępuje niektóre z wątłych i wątpliwych cech teorii Grice'a, przedstawiają Sperber i Wilson (1986/2011).

[61] Jakie mam prawo mówić o opiniach i życzeniach tych nie do końca jeszcze świadomych przodków? Moja teoria przekonań i pragnień zaprezentowana w *The Intentional Stance* broni poglądu, że nie istnieje żaden powód, aby umieścić te pojęcia w cudzysłowie: zachowanie „niższych" zwierząt (nawet żab) jest tak samo dobrą dziedziną wyjaśnienia w nastawieniu intencjonalnym, z przypisywanymi w jej ramach przekonaniami i pragnieniami, jak zachowanie istot ludzkich. Jednak czytelnicy, którzy nie zgadzają się z tą teorią, mogą rozumieć te terminy w metaforycznie rozszerzonym sensie.

[62] Więcej o nadal nierozwiązanej empirycznej kwestii, czy małpy człekokształtne i zwykłe są zdolne do umyślnego oszustwa, piszą Dennett (1983, 1988c, 1988d, 1989a); Whiten (1988); Whiten i Byrne (1988).

[63] W *Ogrodzie o rozwidlających się ścieżkach* Jorge Luis Borges (1972) stworzył diabelsko sprytną wersję tej strategii, jednak powstrzymam się od jej przedstawienia, gdyż nie chcę zdradzać fantastycznego zakończenia.

[64] Ten podrozdział opiera się na moim artykule *Memes and the Exploitation of Imagination* (1990a).

[65] Jeśli interesuje cię spór wokół ewolucji języka, zobacz Pinker i Bloom (1990) oraz późniejsze komentarze.

[66] Tytuł angielskiej piosenki ludowej (przyp. tłum.).

[67] Jest to marsz z popularnej operetki *Mikado* W.S. Gilberta z librettem A. Sullivana z roku 1885. W polskim przekładzie Adolfa Kitschmana z 1888 roku tekst *Behold the Lord High Executioner* brzmiałby znacznie nawet na uroczystości państwowej: „Kto żyw więc krzycz: Zdrowia mu każdy chętnie życz" (przyp. red. nauk.).

[68] Puryści mogą zaprotestować przeciwko takiemu użyciu pojęcia *maszyna wirtualna*, w sensie nieco szerszym niż obowiązujący w informatyce. Odpowiadam: kiedy widzę wygodny element, to jak Matka Natura dokonuję jego „egzaptacji" i używam go w szerszym kontekście (Gould 1980).

[69] A może to *w ogóle* nie być maszyna wirtualna. Może to być szyta na miarę, sprzętowa i realna wyspecjalizowana maszyna, na przykład maszyna Lispa, która pochodzi od maszyn *wirtualnych* Lispa i która jest zaprojektowana aż do poziomu samych układów scalonych, mających obsługiwać język programowania Lisp.

[70] „Logiczne neurony" McCullocha i Pittsa (1943) zostały tak naprawdę stworzone

jednocześnie z wynalezieniem komputera szeregowego i wpłynęły na myślenie von Neumannowskie, co z kolei doprowadziło do powstania perceptronów w latach pięćdziesiątych, przodków dzisiejszego koneksjonizmu. Papert (1988) krótko relacjonuje tę historię.

[71] Więcej o następstwach rzeczywistej prędkości i jej efektach dla sztucznej inteligencji piszę w rozdziale *Fast Thinking* w książce *The Intentional Stance* (1987a).

[72] Interesującą dyskusję dotyczącą (pozornej) niezgodności między tymi dwiema szkołami myśli w dziedzinie sztucznej inteligencji, rozumowanie kontra poszukiwanie, znajdziesz w Simon i Kaplan 1989, s. 18–19.

[73] Błędy językowe polegające na zamienieniu miejscami dwóch początkowych głosek następujących po sobie wyrazów (przyp. tłum.).

[74] Dan Sperber i Deirdre Wilson (1986/2011) otwierają nową perspektywę badań nad komunikacją, wymagającą modeli tego, jak w rzeczywistości coś *funkcjonuje*, w nadawcy i w odbiorcy, w przeciwieństwie do praktyki ostatnio panującej wśród filozofów i językoznawców, nieprzywiązujących wagi do mechanizmów, ale odwołujących się do racjonalnych rekonstrukcji domniemanych zadań i ich wymagań. Pozwala to Sperberowi i Wilson wziąć pod uwagę praktyczność i wydajność: zasadę najmniejszego wysiłku oraz zagadnienia związane z tempem działania i prawdopodobieństwem. Z tej nowej perspektywy pokazują, jak pewne tradycyjne „problemy" znikają – szczególnie problem tego, jak odbiorca odnajduje „trafną" interpretację intencji nadawcy. Mimo że nie przedstawiają swojego modelu na poziomie procesów ewolucyjnych, o których właśnie mówiliśmy, z pewnością zaprasza on do takiego rozwinięcia.

[75] Jak pisze Levelt (1989, s. 16): „Gdyby dało się pokazać na przykład, że na generowanie wiadomości bezpośrednio wpływa dostępność tematów czy form słów, moglibyśmy mieć dowód na bezpośrednią informację zwrotną z Formulatora do Konceptualizatora. Jest to pytanie empiryczne i jest możliwe jego przetestowanie. [...] Jak na razie świadectwa na rzecz hipotezy takiej informacji zwrotnej są negatywne". Świadectwa, których przeglądu dokonuje, pochodzą ze ściśle kontrolowanych eksperymentów, w których nadawca otrzymał bardzo konkretne zadanie: na przykład *opisanie obrazka na ekranie tak szybko, jak tylko się da* (s. 276–282). Ogólnie rzecz biorąc, stanowi to świadectwo negatywne – przynajmniej ja byłem zaskoczony brakiem efektu w tych eksperymentach – jednak, jak mówi Levelt, nie jest to rozstrzygające. Nie będzie twierdzeniem *ad hoc*, że sztuczność tych sytuacji eksperymentalnych zagłuszyła oportunistyczny/twórczy wymiar użycia języka. Być może Levelt ma jednak rację; niewykluczone, że jedyna informacja zwrotna z Formulatora do Konceptualizatora jest *niebezpośrednia*: jest rodzajem informacji zwrotnej, którą ktoś mógłby stworzyć *jedynie* przez mówienie tylko i wyłącznie do siebie, a potem tworząc opinię na temat tego, co uważa, że sam powiedział.

[76] Levelt powiedział mi, że sam ma zwyczaj polowania na dowcipy słowne (jest rodzimym użytkownikiem holenderskiego) i dokładnie wie, jak to robi: „Praktyka, którą rozwijam, odkąd pamiętam, to przekręcanie niemalże każdego słowa, które usłyszę. Następnie (dosyć świadomie) sprawdzam rezultaty dotyczące znaczenia. W 99,9 procenta przypadków nie wychodzi z tego nic śmiesznego. Jednak jeden przypadek na tysiąc jest doskonały, i to właśnie jego natychmiast wyrażam" (osobisty kontakt). Jest to idealny przykład von neumannowskiego rozwiązywania problemów: szeregowego, kontrolowanego – *i świadomego*! Nie wiemy jednak, czy istnieją inne, bardziej pandemoniczne, sposoby nieświadomego tworzenia dowcipów.

[77] Według *Oxford Dictionary of Quotations* (wydanie II, 1953) to słynne zdanie jest również przypisywane Phineasowi T. Barnumowi. Barnum był znamienitym absolwentem i hojnym ofiarodawcą mojej uczelni, czuję się więc zobligowany, aby zwrócić uwagę na

możliwość, że to nie Lincoln był twórcą tego wysoce replikującego się memu.

[78] Przypomina to pogląd Freuda na temat funkcjonowania „przedświadomości": „Pytanie: [...] Jak coś staje się przedświadome? A odpowiedź na nie brzmiałaby: Wskutek pojawienia się związku z odpowiednimi wyobrażeniami słownymi" (Freud 1921/2012, s. 64–65).

[79] Dowiedziałem się od Levelta, że badania, które trwają w Max Planck Institute for Psycholinguistics w Nijmegen, rodzą pewne wątpliwości co do tego. Praca Heeschera sugeruje, że na pewnym poziomie osoby cierpiące na afazję żargonową czują stres związany z upośledzeniem i wydaje się, że przyjmują strategię powtarzania, mając nadzieję na porozumienie komunikacyjne.

[80] Kolejnym nietypowym zjawiskiem językowym jest znany symptom schizofrenii: „słyszenie głosów". Obecnie naukowcy zdecydowanie przyjmują, że głos, który schizofrenik „słyszy", należy do niego; mówi po cichu do siebie, nie zdając sobie z tego sprawy. Przeszkoda tak banalna, jak nakazanie pacjentowi szerokiego otwarcia ust wystarczy, aby głosy ustały (Bick i Kinsbourne 1987). Zobacz również Hoffman (1986) i komentarz Akins i Dennetta *Who May I Say Is Calling?* (1986).

[81] Edelman (1989) jest jedynym teoretykiem, który próbował połączyć to w całość: od szczegółów neuroanatomicznych przez psychologię poznawczą po najbardziej zawiłe spory filozoficzne. Rezultatem jest pouczająca porażka. Pokazuje dokładnie, ile rodzajów pytań musi uzyskać odpowiedź, zanim będziemy mogli twierdzić, że stworzyliśmy całościową teorię świadomości, ale pokazuje też, że żaden teoretyk nie potrafi docenić wszystkich subtelności problemów, którymi zajmują się różne dziedziny. Edelman przeinaczył, a następnie ostro skrytykował prace swoich wielu potencjalnych sprzymierzeńców, przez co jego teoria nie może już liczyć na empatyczne odczytania kompetentnych czytelników, a tego przecież potrzebuje, jeśli ma zostać skorygowana poprzez naprawienie błędów i niedociągnięć. Pojawia się tu podobna możliwość, że i ja nie doceniam niektórych dokonań, z którymi się na tych stronach nie zgadzam. Ba, jestem tego pewien i mam nadzieję, że autorzy prac, które tu błędnie zinterpretowałem, postarają się jeszcze raz wyjaśnić mi to, czego nie zdołałem pojąć.

[82] Funkcjonaliści mają nawyk „pudełkologii" – rysowania wykresów, które przedstawiają składowe funkcje w osobnych „pudełkach", jednocześnie wyraźnie zaprzeczając, że owe pudełka mają znaczenie anatomiczne. (Ja sam nie jestem bez winy, bo robiłem to i podżegałem innych; zob. ryciny w *Brainstorms*, rozdz. 7, 9 i 11). Nadal uważam, że „w zasadzie" jest to dobra taktyka, jednak w praktyce zwykle zaślepia funkcjonalistów na konkurencyjne dekompozycje funkcji, a szczególnie na możliwość wielokrotnych, nałożonych na siebie funkcji. Obraz przestrzennej separacji między pamięcią roboczą a pamięcią długotrwałą – obraz stary jak ptaszarnia Platona – odgrywa niebanalną rolę w rozumieniu zadań poznawczych przez teoretyków. Trafny przykład: „Potrzeba symboli pojawia się, gdyż nie jest możliwe, aby wszystkie struktury zaangażowane w obliczenia zaistniały wcześniej w miejscu fizycznym, gdzie odbywa się obliczanie. Jest więc konieczne sięgnięcie do innych (dalszych) części pamięci po dodatkową strukturę" (Newell, Rosenbloom i Laird 1989, s. 105). Prowadzi to dosyć bezpośrednio do obrazu *ruchomych symboli*, następnie (w przypadku tych, którzy bezkrytycznie popierają ten obraz) do sceptycyzmu wobec wszelkiej architektury koneksjonistycznej, gdyż elementy w takiej architekturze, które są najbliższe symbolom – węzły w taki czy inny sposób kotwiczące semantykę systemu – są nieruchome w sieci połączeń. Zob. np. Fodor i Pylyshyn (1988). Ten problem stałych i ruchomych elementów semantycznych jest jednym ze sposobów ujęcia ważnego, a nierozwiązanego problemu kognitywistyki. Nie jest to prawdopodobnie dobre ujęcie, ale nie zniknie, dopóki nie zostanie zastąpione lepszą wizją, nie historycznie odrzucaną, lecz dobrze podpartą przez fakty neuroanatomii funkcjonalnej.

[83] Innymi słowy, uroki „przyczynowej teorii odniesienia” są tak samo oczywiste dla kognitywistów jak dla filozofów.

[84] Fodor mówi o odmianie tego problemu w analizach dotyczących „pomyślenia pojęcia” (1990, s. 80–81).

[85] Teorie reflektorów są popularne od lat. Prymitywne teorie popełniają błąd polegający na zbyt dosłownym założeniu, że to, co reflektor różnorako podświetla czy uwydatnia w danym momencie, to region *przestrzeni wizualnej* – tak jak reflektor w teatrze podświetla tylko jeden obszar sceny. Łatwiejsze do obronienia – choć w tym przypadku również bardziej impresjonistyczne – teorie reflektora zakładają istnienie... uwydatnionych *konceptualnych* czy *semantycznych przestrzeni* (jeśli potrafisz, to wyobraź sobie reflektor teatralny, który może oświetlić tylko Kapuletów lub wszystkich i tylko kochanków). Allport (1989) przedstawia trudności związane z teoriami reflektora.

[86] Pokrewnymi rozróżnieniami są moja trójca nastawienia intencjonalnego, nastawienia konstrukcyjnego i nastawienia fizycznego (Dennett 1971) oraz wskazanie „poziomu wiedzy” ponad „poziomem fizycznego systemu symboli” u Allena Newella (1982). Zob. Dennett (1987a, 1988e) oraz Newell (1988).

[87] Jak napisał Marr (1982, s. 19): „Dzięki podzieleniu wyjaśnień na trzy poziomy możliwe staje się jasne stwierdzenie, co jest obliczane i dlaczego, oraz konstruowanie teorii stwierdzających, że to, co jest obliczane, jest w pewnym sensie optymalne lub że poprawne funkcjonowanie tego, co obliczane, jest zagwarantowane”. Zob. Dennett (1971, 1983, 1987a, 1988d) oraz Ramachandran (1985) w sprawie zysków i strat związanych z tego rodzaju inżynierią odwrotną.

[88] W *Minimal Rationality* (1986) filozof Christopher Cherniak analizuje szanse i ograniczenia procesów dedukcyjnych z rodzaju tych, które stają się możliwe dzięki istnieniu otwartego forum. Zob. też Stalnaker (1984).

[89] Jackendoff (1987) przyjmuje trochę inną taktykę. Dzieli problem umysł-ciało na dwie części i przedstawiając swoją teorię, odpowiada na pytanie, jak *obliczeniowy* umysł pasuje do ciała; pozostawia go to z nierozwiązanym problemem „umysł-umysł” – jaka jest relacja między umysłem fenomenologicznym i obliczeniowym. Zamiast zgadzać się, że ta tajemnica pozostała, chcę pokazać, jak model wielokrotnych szkiców, wspólnie z metodą heterofenomenologii, rozwiązuje oba problemy jednocześnie.

[90] To przydatne ułatwienie to jeden z wielu drobiazgów, które później były eksplorowane przez badaczy, i obecnie istnieje sporo dowodów na efekty „bezwładu” i „pędu” w transformacjach obrazów. Zob. Freyd (1989).

[91] Imponująca, ale wyraźnie szarpana animacja w popularnym programie IBM PC *Flight Simulator* pokazuje ograniczenia animacji w czasie rzeczywistym całkiem złożonych scen trójwymiarowych na małym komputerze.

[92] Nazywam to urządzenie *Vorsetzer*, ponieważ przypomina mi o wspaniałym niemieckim urządzeniu o tej samej nazwie, grającym na fortepianie, na które to urządzenie składała się oddzielna jednostka 88 „palców”, mogąca „usiąść przed” zwykłym fortepianem oraz naciskać klawisze i pedały od zewnątrz, jak pianista-człowiek. (Ważne, aby nie zapomnieć, że to urządzenie to *Vorsetzer* – „siadający przed” – a nie *Vorsetzer* – prezes czy prezydent!)

[93] Gdy mamy już etykiety, możemy opowiedzieć o *każdej* właściwości obiektu, nie tylko o właściwościach przestrzennych czy wizualnych – jak w starym dowcipie z kolorowanki: „Oto mój szef. Pokoloruj go odpychająco”.

[94] W Kosslyn (1980) zobacz dyskusję o formacie. Jackendoff (1989) podobnie analizuje to, co nazywa „formą struktur informacyjnych”.

[95] Kosslyn (1980) nie tylko szczegółowo broni swojego szczególnego zestawu odpowiedzi na te pytania (w owym czasie), ale również daje wspaniały przegląd innych eksperymentalnych i teoretycznych badań w zakresie wyobrażeń. Dobre podsumowanie tych prac w kolejnej dekadzie można znaleźć w Farah (1988) oraz w Finkle, Pinker i Farah (1989).

[96] Jak powiada Marvin Minsky (1985, s. 151): „Nie ma nic nadzwyczajnego w idei wyczuwania zdarzeń w mózgu. Agenty to agenty – i jest tak samo łatwo agentowi mieć wbudowane elementy wykrywające *spowodowane przez mózg zdarzenie mózgowe*, jak wykrywać *spowodowane przez świat zdarzenie mózgowe*".

[97] W rozdziale 7 (s. 313) zapytałem: „zwabiać na powierzchnię *czego*?" i obiecałem, że odpowiem na to pytanie później. Oto moja odpowiedź. (Metaforyczna) powierzchnia jest zdeterminowana przez format interakcji między częściami.

[98] Ciekawe jest porównanie różnych śladów pomysłu użytkownika w mózgu w pracy zasadniczo odmiennych myślicieli. Oto Minsky (1985, s. 75, 59): „Wyolbrzymiając nieco tę kwestię, na to, co nazywamy »świadomością«, składa się zaledwie lista błyskająca od czasu do czasu na ekranie umysłu, z której korzystają inne systemy. [...] Podziel mózg na dwie części, A i B. Połącz dane wejściowe i wyjściowe mózgu A ze światem rzeczywistym – tak aby mógł wyczuwać, co się tam dzieje. Jednak w ogóle nie łącz mózgu B ze światem zewnętrznym; zamiast tego połącz je tak, że mózg A to świat mózgu B!". Minsky mądrze powstrzymuje się od sugerowania linii dzielącej mózg anatomicznie na dwie części, ale niektórzy są na to gotowi. Gdy Kosslyn po raz pierwszy rozważał świadomość jako maszynę wirtualną, skłaniał się ku ulokowaniu użytkownika w płacie czołowym (zob. też Kosslyn 1980, s. 21), a ostatnio Edelman sam argumentował na rzecz tego samego wniosku, wyrażanego w kategoriach „wartościującej pamięci własnej/niewłasnej", którą umieścił w płacie czołowym i przypisał jej zadanie interpretowania wytworów pozostałych części mózgu (Edelman 1989, od s. 102).

[99] Jest to wątek co najmniej bardzo spokrewniony z kluczowym motywem późniejszej twórczości Wittgensteina, ale nie chciał on podać jakiegoś pozytywnego opisu czy modelu relacji między tym, co mówimy, a tym, o czym mówimy, gdy (pozornie) relacjonujemy nasze stany umysłowe. Filozofka Elizabeth Anscombe w swoim frustrująco niejasnym, klasycznym dziele *Intention* (1957) próbowała wypełnić tę lukę pozostawioną przez Wittgensteina, twierdząc, że błędem jest uważanie, iż *wiemy*, jakie są nasze intencje; możemy raczej *powiedzieć*, jakie są nasze intencje. Podjęła również próbę scharakteryzowania kategorii rzeczy, które możemy *wiedzieć bez obserwacji*. Niedoskonałe rozważania na ten temat snuję w mojej książce *Content and Consciousness* (1969), w rozdziałach 8 i 9. Zawsze myślałem, że jest coś słusznego, ważnego i oryginalnego w tych fragmentach. Drugie moje do niego podejście można znaleźć w *Toward a Cognitive Theory of Consciousness* (1978), artykule później przedrukowanym w *Brainstorms*, szczególnie w częściach IV i V (s. 164–171). Ten podrozdział to moja aktualna próba rozjaśnienia tych idei i odejścia od dwóch wcześniejszych wersji.

[100] W *Brainstorms* wykorzystałem tę cechę psychologii potocznej w moich rozważaniach o „β-rozmaitościach w przekonaniach fenomenologicznych" (1978a, od s. 177).

[101] Tim Shallice w książce *From Neuropsychology to Mental Structure* (1988) przedstawia aktualne i dobrze uargumentowane przemyślenia dotyczące rozumowania związanego z analizowaniem tych eksperymentów natury. Kilka niedawno wydanych książek przytacza dobre, znane relacje niektórych z tych fascynujących przypadków: *The Shattered Mind* (1975) Howarda Gardnera oraz *Mężczyzna, który pomylił swoją żonę z kapeluszem* (1985/2001) Olivera Sacksa.

[102] Zauważmy, że szczegóły uszkodzeń neurologicznych same w sobie (bez zaprzeczeń) niczego by nie udowodniły; jedynie poprzez połączenie ich z (wiarygodnymi)

relacjami oraz świadectwami behawioralnymi zyskujemy jakąkolwiek hipotezę o tym, które części mózgu są konieczne dla którego świadomego zjawiska.

[103] Bez potwierdzenia w postaci skanów mózgu ukazujących zniszczenia w korze z pewnością rozpowszechniony byłby sceptycyzm co do prawdziwości obszarów ślepych pacjentów ze ślepowidzeniem. Zob. np. Campion, Latto i Smith (1983) oraz Weiskrantz (1988).

[104] Filozof Colin McGinn (1991) opowiada o wyimaginowanym pacjencie cierpiącym na ślepowidzenie: „Behawioralnie może funkcjonować właściwie jak osoba widząca; fenomenologicznie dziwi się, że nie widzi" (s. 111). Jest to po prostu nieprawda; zupełnie nie może funkcjonować behawioralnie jak osoba widząca. McGinn przechodzi do wzmocnienia swojego niezwykłego twierdzenia: „Poza tym, bądźmy przez chwilę naiwni, czy pacjenci ze ślepowidzeniem nie *wyglądają* zupełnie tak, jak gdyby mieli doświadczenia wizualne, gdy przekazują swoje zaskakujące rozróżnienia? [...] Nie wyglądają jak ludzie, którzy *niczego* nie przeżywają" (s. 112). I znów jest to nieprawda. Tak naprawdę wyglądają, jak gdyby *nie* mieli przeżycia wizualnego, *ponieważ muszą dostać sygnał*. Gdyby nie musieli tego sygnału otrzymać, rzeczywiście wyglądaliby, jak gdyby mieli doświadczenia wizualne – tak bardzo, że nie wierzylibyśmy w ich zaprzeczenia!

[105] „Gdyby mógł posłuchać swojej reakcji skórno-galwanicznej, byłby w lepszym stanie" – Larry Weiskrantz komentujący jednego ze swoich pacjentów ze ślepowidzeniem, ZIF, Bielefeld, maj 1990.

[106] Moją odpowiedzią na to pytanie jest moja książka *The Intentional Stance* (1987).

[107] Czy ta identyfikacja jest *efektem* następującym po tym, jak uświadomiła to sobie, czy wcześniejszą *przyczyną* tego, że stała się świadoma? Jest to pytanie – orwellowskie czy stalinowskie? – którego, według modelu wielokrotnych szkiców, nie warto zadawać.

[108] W normalnych warunkach zlokalizowanie (dostrzeżenie go) i identyfikacja idą ze sobą w parze; dostrzeżenie obiektu do zidentyfikowania jest warunkiem tej identyfikacji. Jednak te normalne warunki ukrywają zaskakujący fakt: maszynerie identyfikacyjna oraz lokalizacyjna są w dużej mierze niezależne w mózgu, gdyż są zlokalizowane w różnych obszarach kory (Mishkin, Ungerleider i Macko 1983) i dlatego mogą być wyłączone niezależnie od siebie. Istnieją rzadkie patologie, w których badany może szybko zidentyfikować, *co* widzi bez żadnej możliwości zlokalizowania tego w przestrzeni prywatnej, oraz odpowiadające im patologie, w których badany potrafi zlokalizować bodziec wizualny – na przykład na niego wskazać – jednocześnie nie dając sobie rady z jego identyfikacją, mimo że w innych przypadkach jego wzrok działa zupełnie normalnie. Psycholog Anna Treisman (1989; Treisman i Gelade 1980; Treisman i Sato 1990; Treisman i Souther 1985) przeprowadziła ważną serię eksperymentów wspierających jej twierdzenie, że *widzenie* powinno być odróżnione od *identyfikacji*. Gdy coś jest widziane, według jej modelu, mózg ustala „żeton" tego obiektu. Żetony to „osobne, czasowe, epizodyczne reprezentacje" – a ich tworzenie jest wstępem do dalszej identyfikacji, czyli czegoś, co zostaje osiągnięte przez przeszukanie pamięci semantycznej za pośrednictwem procesu w rodzaju tych, które modelują systemy produkcyjne. Żeton nie musi być jednak zdefiniowany przez konkretną lokalizację w przestrzeni prywatnej, jeśli rozumiem jej model, a z tego powodu nie jest wykluczone, że osoby w stanie, w jakim znajdowała się Betsy (zanim odnalazła naparstek), mogłyby osiągnąć wyniki lepsze niż przypadkowe, jeśli zostałby na nich wymuszony wybór dotyczący tego, czy naparstek był obecnie w ich polu widzenia, czy nie. Eksperymenty z tym związane znajdziesz w: Pollatsek, Rayner i Henderson (1990).

[109] Na przykład opóźnienie reakcji na niektóre z tych zadań percepcyjnych, nawet u wyszkolonych osób badanych, jest dość długie – osiem do dziesięciu sekund dla różnych prostych identyfikacji (Bach-y-Rita 1972, s. 103). Pokazuje to, że przepływ informacji

w widzeniu protetycznym jest niezwykle powolny w porównaniu z widzeniem normalnym.

[110] „Szybkość transmisji w bodach" jest standardowym określeniem prędkości przepływu informacji cyfrowych w bodach (oznacza to mniej więcej: bity na sekundę). Jeśli na przykład komputer komunikuje się z innymi komputerami przez sieć telefoniczną, może transmitować ciągi bitów z prędkością 1200 lub 2400 bodów albo o wiele szybciej. Wymagana jest czterokrotnie większa szybkość transmisji w stosunku do transmisji animacji w wysokiej rozdzielczości w czasie rzeczywistym – jasny przykład, w którym obraz jest rzeczywiście wart więcej niż tysiąc słów. Zwykle sygnały telewizyjne są analogowe, jak płyta gramofonowa, a nie cyfrowe, jak CD, więc prędkość ich przepływu jest określana mianem *przepustowości*, a nie szybkości transmisji w bodach. Pojęcie powstało przed komputerami; kod Baudot, nazwany nazwiskiem swojego wynalazcy (jak alfabet Morse'a), był standardowym, międzynarodowym kodem telegraficznym przyjętym w roku 1880, a szybkość transmisji w bodach była liczbą elementów kodu transmitowanych na sekundę. Używając pojęcia „szybkość transmisji w bodach", a nie „przepustowość", nie zamierzam sugerować, że przetwarzanie informacji przez mózg lepiej opisywać w kategoriach cyfrowych.

[111] Istnieją inne rodzaje algorytmów kompresji, które nie polegają na dzieleniu obrazu na obszary o tym samym kolorze w opisany powyżej sposób, ale nie będę się nimi zajmował.

[112] Inne stworzenia mają inne bryły barw – albo hiperbryły! My widzimy „trójchromatycznie": mamy trzy rodzaje białka posiadającego zdolność pochłaniania światła w czopkach. Inne gatunki, takie jak gołębie, widzą „tetrachromatycznie"; ich subiektywna przestrzeń barw musiałaby być zaprezentowana, liczbowo, jako czterowymiarowa hiperprzestrzeń. Inne gatunki mają widzenie dwuchromatyczne, a ich rozróżnienia kolorów mogłyby zostać odwzorowane na dwuwymiarowej płaszczyźnie. (Zwróćmy uwagę, że „czerń i biel" to tylko jednowymiarowy schemat reprezentacyjny, a wszystkie możliwe szarości są przedstawiane jako różne odległości na linii pomiędzy 0 i 1). Przemyślenia na temat następstw owej niewspółmierności systemów barw znajdziesz w Hardin (1988) oraz Thompson, Palacios i Varela (1992).

[113] Niejako w odpowiedzi na tę sugestię V.S. Ramachandran i R.L. Gregory (1991) przeprowadzili eksperymenty z czymś, co nazywają (uważam, że zwodniczo) „sztucznie wywołanym obszarem ślepym", w którym znaleźli silne świadectwa mówiące o stopniowym wypełnianiu tekstur i szczegółów. Nie ma fundamentalnej różnicy między ich warunkami eksperymentalnymi a tymi opisanymi przeze mnie; w ich eksperymentach jest konkurencja między dwoma źródłami informacji, z których jedno zostaje odrzucone (stopniowo). Zjawisko stopniowego wypełniania przestrzeni tekstur jest ważnym odkryciem, lecz nie prowadzi nas dalej niż model w duchu ryciny 11.11. Dalsze kwestie dotyczące tych eksperymentów muszą zostać wyjaśnione, zanim ich interpretacja będzie ostateczna.

[114] Na przykład pierwsze eksperymenty Rogera Sheparda z umysłowym obracaniem rysunków brył pokazały, że osobom badanym z pewnością *wydawało się*, że pojawiają się w nich z grubsza ciągłe, obracające się reprezentacje kształtów, które sobie wyobrażali, lecz potrzeba było kolejnych eksperymentów sprawdzających rzeczywiste czasowe właściwości odnośnych reprezentacji, aby (częściowo) potwierdzić hipotezę, że rzeczywiście robili to, co im się wydawało, że robili. (Zob. Shepard i Cooper 1982).

[115] W załączniku B zaproponuję pewne „eksperymenty z tapetą", które mogą osłabić to empiryczne twierdzenie.

[116] Dla kontrastu zobacz Bisiach i in. (1986) oraz McGlynn i Schacter (1989), których modele anosognozji są podobne, ale opierają się na „pudełkowaniu" oddzielnych *systemów*, szczególnie u McGlynn i Schactera, którzy zakładają istnienie *systemu świadomej przytomności*,

czerpiącej informacje z modułów.

[117] Wariacje zbudowane na tym motywie można znaleźć w Humphrey (1976, 1983a) oraz Thompson, Palacios i Varela (1991).

[118] Obecnie filozofowie bardzo lubią pojęcie *rodzajów naturalnych*, przywróconych w filozofii przez Quine'a, który może teraz żałować sposobu, w jaki stało się ono substytutem wątpliwego, choć skrycie popularnego pojęcia *esencji*. „Zielone rzeczy, a przynajmniej zielone szmaragdy są rodzajem", zauważa Quine (1969, s. 116), ujawniając, że docenia fakt, iż podczas gdy szmaragdy mogą być rodzajem naturalnym, *zielone* rzeczy prawdopodobnie nim nie są. Obecne rozważania mają na celu zapobiegnięcie jednemu z kuszących błędów naturalizmu kanapowego: założeniu, że wszystko, co tworzy natura, jest rodzajem naturalnym. Barwy *nie* są „rodzajami naturalnymi" właśnie *dlatego*, że nie są wytworem ewolucji biologicznej, która tworząc rodzaje, ma tolerancję na niechlujne granice, mogące przerazić każdego filozofa z tendencją do dobrych, jasnych definicji. Gdyby życie jakiegoś stworzenia zależało od wrzucenia do jednego wora księżyca, sera pleśniowego i rowerów, to można mieć pewność, że Matka Natura znalazłaby sposób, aby „widzieć" je jako „intuicyjnie ten sam rodzaj rzeczy".

[119] Prymatolożka Sue Savage-Rumbaugh poinformowała mnie, że bonobo, czyli szympanse karłowate, dorastające w laboratorium nie okazują żadnej wrodzonej niechęci do węży, w przeciwieństwie do szympansów.

[120] Ważna byłaby nagłość, bo gdyby stało się to stopniowo, być może byłoby niezauważalne. Jak wskazał Hardin (1990), stopniowe żółcenie się twoich soczewek z wiekiem powoli przesuwa wyczucie barw podstawowych; gdyby pokazano ci koło barw i poproszono o wskazanie na czystą czerwień (czerwień bez domieszki pomarańczu czy fioletu), wówczas to, gdzie na kontinuum wskażesz, jest częściowo cechą twojego wieku.

[121] „Stąd, że tak chętnie powiedzielibyśmy: »Ważne jest *to*« – sami sobie wskazując na doznanie – widać już, jak wielką mamy skłonność, by powiedzieć coś, co nie jest przekazem informacji", Wittgenstein (1953/2000), i298, s. 146.

[122] Byłoby aktem desperackiej, intelektualnej nieuczciwości cytowanie tego fragmentu bez kontekstu!

[123] Zob. także Lockwood (1989, s. 15–16): „Jak odczuwalibyśmy świadomość, gdyby *sprawiała wrażenie* miliardów maleńkich atomów poruszających się w miejscu?".

[124] Główne idee zaprezentowałem w moich rozmyślaniach o Borgesie w *The Mind's I* (Hofstadter i Dennett 1981, s. 348–352) i spiąłem je w całość podczas odczytu *The Self as the Center of Narrative Gravity* (*Jaźń jako środek narracyjnej ciężkości*) zaprezentowanego na Houston Symposium w 1983 roku. Gdy czekałem na pojawienie się artykułów z sympozjum drukiem, opublikowałem dosyć skróconą wersję mojego wystąpienia w „Times Literary Supplement" z września 1988, pod nudnym tytułem – nie moim – *Dlaczego każdy jest powieściopisarzem*. Wersja oryginalna, zatytułowana *The Self as the Center of Narrative Gravity*, ukazała się w 1992 roku w tomie *Self and Consciousness* pod redakcją F. Kessela, P. Cole'a i D. Johnsona, Erlbaum, Hillsdale, NJ.

[125] Przekład poprawiony (przyp. red. nauk.).

[126] Co ciekawe, Nagel w 1971 roku wyczerpująco odpowiadał już na to pytanie (1971/1997, s. 185), zanim zainteresował się nietoperzami – tematem, którym zajmiemy się w kolejnym rozdziale.

[127] A skąd wiemy, że *my* coś robimy? Skąd bierzemy początkową wiedzę o samym sobie, która jest nam w tym przypadku potrzebna? Wydaje się to naprawdę fundamentalnym pytaniem dla niektórych filozofów (Castañeda 1967, 1968; Lewis 1979; Perry 1979) i zrodziło literaturę o nieprześcignionej zawiłości. Jeśli jest to istotny problem filozoficzny, musi być coś

nie tak z „banalną” odpowiedzią (ale nie widzę co): otrzymujemy podstawową, pierwotną wiedzę o samych sobie tak samo jak homary; po prostu jesteśmy tak skonstruowani.

[128] Spróbuj wyobrazić sobie stan umysłu Ji Hu-Mina, mojego studenta z Pekinu, dla którego wstępem do angloamerykańskiej filozofii umysłu (podczas gdy jego angielski nadal był dosyć elementarny) był udział w seminarium, gdzie studenci i profesorowie żywo debatowali o tym, co by się stało, gdyby cała populacja Chin została w jakiś sposób zmuszona do uczestnictwa w ogromnej realizacji przypuszczalnie świadomego programu AI (przykład Blocka), a następnie przeszli, z taką samą nieświadomością wrażliwości chińskiego obserwatora, do chińskiego pokoju Searle’a.

[129] Ostateczne obalenie tego eksperymentu, do którego Searle nadal adekwatnie się nie odniósł, zostało zaprezentowane przez Hofstadtera w Hofstadter i Dennett (1981), na stronach 373–382. W późniejszych latach pojawiła się fala zdecydowanej krytyki. W *Fast Thinking* (w: Dennett 1987a) przedstawiłem nową diagnozę źródeł nieporozumienia w tym eksperymencie. Odpowiedź Searle’a to deklaracja, niepoparta żadnym argumentem, że wszystkie zagadnienia były bez związku z eksperymentem (Searle 1988b). Żaden magik nie lubi, gdy jego sztuczki zostają wyjaśnione publiczności.

[130] Autor dziękuje za uwagi do wcześniejszej wersji tego tekstu, które otrzymał od Mateusza Hohola, Tomasza Korbaka, Pawła Gładziejewskiego i Miry Marcinów.