

# WYMYŚLEO

CHALMERS

świadomy  
umysł

WYDAWNICTWO NAUKOWE PWN

**Świadomy umysł** należy do najgłośniejszych książek z filozofii umysłu ostatnich kilkunastu lat. Przedmiotem analizy autora jest świadome przeżywanie – świadomość fenomenalna. Chalmers uważa, że dotąd badano świadomość tylko pod kątem „łatwych problemów”. Za łatwe zaś uznaje te, które można rozwiązać dzięki współczesnej nauce, czyli zagadnienia związane ze sposobem przetwarzania informacji, architekturą umysłu i innymi jego funkcjonalno-relacyjnymi własnościami. „Trudne problemy” pojawiają się natomiast wtedy, gdy próbuje się wyjaśnić owo nieusuwalne subiektywne poczucie, jak to jest być świadomym. Tu, zdaniem Chalmersa, nie pomoże nauka: na gruncie obecnie znanych koncepcji nie można zrozumieć, z czego wynika to poczucie i na czym polega. Sam więc bierze się za analizę i proponuje naturalistyczną teorię świadomego przeżywania oraz swego rodzaju antyredukcyjny funkcjonalizm. Wykorzystuje przy tym swoją rozległą wiedzę naukową ze współczesnej fizyki, zwłaszcza mechaniki kwantowej, informacji czy sztucznej inteligencji.

**David Chalmers** (ur. 1966) – profesor Australijskiego Uniwersytetu Narodowego. Wychowanek uniwersytetów amerykańskich, uczestnik programu Filozofia–Neuro-nauka–Psychologia na uniwersytecie w St. Louis. Autor książek i artykułów z zakresu filozofii umysłu, szczególnie problemów związanych ze świadomością. Jego głównym dziełem jest prezentowany właśnie czytelnikowi **Świadomy umysł**. Chalmers to barwna postać świata akademickiego. Jest niezwykle aktywny i otwarty na nowe sposoby uprawiania filozofii, uczestniczy w imponującej liczbie konferencji na całym świecie, także w Polsce. To autor z wyobraźnią i charakterystycznym poczuciem humoru – i nawet niedociągnięcia w jego koncepcji są pouczające.



MY  
BEST  
BOOK

David  
**CHALMERS**

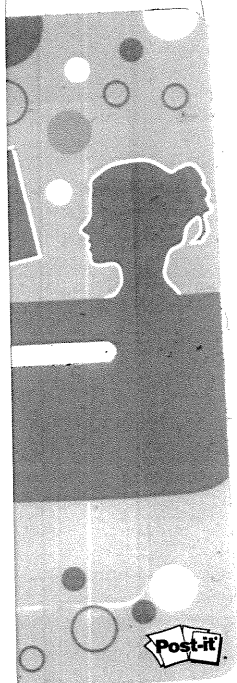
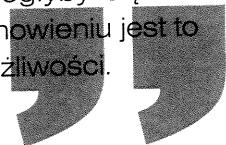
**Świadomy  
umysł**  
w poszukiwaniu  
teorii fundamentalnej

przekład  
Marcin Miłkowski



WYDAWNICTWO NAUKOWE PWN  
WARSZAWA 2010

„Gdy schodzimy na skali od ryb i ślimaków przez proste sieci neuronowe aż do termostatów, w którym momencie świadomość powinna zaniknąć? Fenomenologia ryb i ślimaków prawdopodobnie nie będzie prymitywna, ale stosunkowo złożona, odzwierciedlając rozróżnienia, które mogą one poczynić. Zanim fenomenologia całkowicie zaniknie, przypuszczalnie stanie się jakąś maksymalnie prostą fenomenologią. Wydaje mi się, że najnaturalniejszym miejscem, gdzie może to nastąpić, jest system z odpowiednio prostą „psychologią postrzeżenia”, taki jak termostat. Wydaje się, że termostat realizuje rodzaj przetwarzania informacji obecny u ślimaka lub ryby, ale sprowadzony do jego najprostszej formy, więc być może mógłby również mieć odpowiedni rodzaj fenomenologii w jej najprostszej postaci. Czyni jedno lub dwa istotne rozróżnienia, od których zależy jego działanie; mnie przynajmniej nie wydaje się nierozsądne, że mogłyby się z tym wiązać rozróżnienia przeżyć. Oczywiście, sprawy mogłyby potoczyć się inaczej, gdy schodzimy w dół na skali złożoności – i nie jest to żaden dowód, że termostaty muszą mieć przeżycia. Wydaje się jednak rozsądne, że sprawy mogłyby się tak właśnie potoczyć, a po zastanowieniu jest to równie naturalne, jak inne możliwości.



# Świadomy umysł

Dane oryginału  
David J. Chalmers

**The Conscious Mind. In Search of a Fundamental Theory**

First Edition

Copyright © 1996 by David J. Chalmers

**The Conscious Mind. In Search of a Fundamental Theory**

First Edition, was originally published in English in 1996. This translation is published by arrangement with Oxford University Press.

Pierwsze wydanie **The Conscious Mind. In Search of a Fundamental Theory** ukazało się w języku angielskim w 1996 roku. Polskie tłumaczenie ukazuje się na podstawie umowy z Oxford University Press.

Projekt okładki i wnętrza  
Agata Muszalska

Redaktor inicjujący  
Monika Michowicz

Redaktor  
Michał Warchala

Rysunki  
s. 37, 174, 383, 514  
Marcin Jażyński

Podręcznik akademicki dotowany przez Ministra Nauki i Szkolnictwa Wyższego

Copyright © for the Polish edition  
by Wydawnictwo Naukowe PWN SA  
Warszawa 2010

ISBN 978-83-01-16498-0

Wydawnictwo Naukowe PWN SA  
02-676 Warszawa, ul. Postępu 18  
tel. 22 69 54 321, faks 22 69 54 288  
e-mail: pwn@pwn.com.pl  
www.pwn.pl

# spis treści

## podziękowania 9

## wstęp: wziąć świadomość poważnie 13

## podstawy 27

### I dwa pojęcia umysłu 29

- 1 czym jest świadomość? 29
- 2 koncepcje umysłu: fenomenalne i psychologiczne 41
- 3 podwójne życie terminów umysłowych 49
- 4 dwa problemy umysł-ciało 61
- 5 dwa pojęcia świadomości 64

### II superweniencja a wyjaśnianie 74

- 1 superweniencja 74
- 2 wyjaśnianie redukcyjne 93
- 3 superweniencja logiczna a wyjaśnianie redukcyjne 101
- 4 prawda pojęciowa a prawda konieczna 108
- 5 prawie wszystko jest logicznie superweniencjne na tym, co fizyczne 139

## nieredukowalność świadomości 169

### III czy świadomość można wyjaśnić redukcyjnie? 171

- 1 czy świadomość jest logicznie superweniencjna na tym, co fizyczne? 171
- 2 fiasko wyjaśniania redukcyjnego 191
- 3 modelowanie poznawcze 198
- 4 wyjaśnianie neurobiologiczne 204

- 5 odwołanie do nowej fizyki **209**
- 6 wyjaśnianie ewolucyjne **212**
- 7 quo vadis, wyjaśnienie redukcyjne? **213**

#### **IV dualizm naturalistyczny 216**

- 1 argument przeciwko materializmowi **216**
- 2 zarzuty oparte na konieczności a posteriori **228**
- 3 inne argumenty za dualizmem **243**
- 4 czy to epifenomenalizm? **260**
- 5 geografia logiczna problematyki **277**
- 6 refleksje na temat dualizmu naturalistycznego **293**

#### **V paradoks sądu fenomenalnego 298**

- 1 świadomość i poznanie **298**
- 2 paradoks sądu fenomenalnego **304**
- 3 o wyjaśnianiu sądów fenomenalnych **316**
- 4 argumenty przeciwko nieistotności eksplanacyjnej **326**
- 5 argument z samowiedzy **327**
- 6 argument z pamięci **340**
- 7 argument z odniesienia **342**

#### **w stronę teorii świadomości 355**

#### **VI spójność między świadomością i poznaniem 357**

- 1 w stronę teorii nieredukcyjnej **357**
- 2 zasady spójności **364**
- 3 więcej o pojęciu przytomności **375**
- 4 eksplanacyjna rola zasad spójności **390**
- 5 spójność jako prawo psychofizyczne **404**

#### **VII nieobecne qualia, blaknące qualia, tańczące qualia 411**

- 1 zasada niezmienności organizacyjnej **411**
- 2 nieobecne qualia **417**
- 3 blaknące qualia **420**

- 4 odwrócone qualia **435**
- 5 tańczące qualia **441**
- 6 funkcjonalizm nieredukcyjny **453**

#### **VIII świadomość i informacja: kilka spekulacji 456**

- 1 w stronę teorii fundamentalnej **456**
- 2 aspekty informacji **458**
- 3 argumenty na rzecz teorii **472**
- 4 czy przeżycia są wszechobecne? **780**
- 5 metafizyka informacji **493**
- 6 pytania otwarte **504**

#### **zastosowania 509**

#### **IX mocna sztuczna inteligencja 511**

- 1 świadomość maszynowa **511**
- 2 o implementacji obliczeń **515**
- 3 w obronie mocnej AI **522**
- 4 chiński pokój i inne zarzuty **525**
- 5 zarzuty zewnętrzne **535**
- 6 wnioski **540**

#### **X interpretacja mechaniki kwantowej 542**

- 1 dwie tajemnice **542**
- 2 ramy mechaniki kwantowej **543**
- 3 interpretacje mechaniki kwantowej **548**
- 4 interpretacja Everetta **561**
- 5 zarzuty wobec interpretacji Everetta **568**
- 6 wnioski **575**

#### **bibliografia 579**

#### **indeks nazwisk 595**

#### **indeks pojęć 601**

## VIII Świadomość i informacja: kilka spekulacji

### <sup>1</sup> w stronę teorii fundamentalnej

Do tej pory wyodrębniliśmy kilka związków między świadomością a procesami fizycznymi zasługujących na miano praw psychofizycznych. Jednym z nich jest zasada spójności łącząca świadomość z przytomnością, czyli dostępnością globalną. Drugim jest bardziej szczegółowa zasada spójności strukturalnej, łącząca strukturę świadomości ze strukturą przytomności. Zasada niezmienności organizacyjnej jest trzecim takim związkiem. Zasady te mogą być *składnikami* ostatecznej teorii świadomości. Umożliwiają nam wykorzystanie faktów fizycznych do przewidywania i wyjaśniania pewnych faktów dotyczących przeżyć świadomych. Z pewnością też *ograniczają* formę ostatecznej teorii świadomości: jeśli taka teoria nie jest zgodna z tymi prawami, jest mało prawdopodobne, aby była prawdziwa. Na tym jednak sprawa się nie kończy. Zasady te nie sumują się do ostatecznej teorii ani niczego, co byłoby do niej zbliżone.

Sęk w tym, że żadna z tych zasad nie jest wiarygodnym potencjalnym *fundamentalnym* prawem w teorii świadomości. Wszystkie wyrażają prawidłowości na dość wysokim poziomie. Na przykład pojęcie przytomności (lub globalnej dostępności) jest pojęciem wysokopoziomowym, a jego granice są nieco nieostre; jest bardzo mało prawdopodobne, by to pojęcie miało występować w prawie fundamentalnym. Zasada niezmienności organizacyjnej jest może mniej nieostra, ale nadal wyraża prawidłowości na

poziomie, który jest daleki od fundamentalnego. Kolejny problem: zasady te zdecydowanie za słabo określają charakter związku psychofizycznego. Wszelkiego rodzaju pytania dotyczące tego związku pozostają bez odpowiedzi. Na przykład, jakiego konkretnie *rodzaju* organizacja wywołuje przeżycia świadome? Jak prosta może być organizacja, zanim przeżycia zanikną? I jak możemy przewidzieć specyfikę przeżycia (nie tylko jego strukturę), znając jego fizyczną podstawę? Chcielibyśmy, żeby kompletna teoria świadomości odpowiedziała na te pytania, ale opisane do tej pory zasady w tym nie pomagają.

Do ostatecznej teorii potrzebujemy zbioru praw psychofizycznych analogicznych do fundamentalnych praw fizyki. Te fundamentalne (lub *podstawowe*) prawa będą ujęte na poziomie łączącym podstawowe własności przeżycia z prostymi cechami świata fizycznego. Prawa powinny być precyzyjne i – wzięte łącznie – nie powinny pozostawiać miejsca na niedookreślenie. W połączeniu z faktami fizycznymi na temat układu powinny umożliwić nam doskonałe przewidywanie faktów fenomenalnych o układzie. Ponadto podobnie jak z podstawowych praw fizyki wynikają wszystkie prawa i prawidłowości wyższego poziomu (przynajmniej w połączeniu z warunkami brzegowymi), podstawowe prawa dotyczące świadomości powinny pociągać za sobą i wyjaśniać rozmaite prawa niepodstawowe, takie jak zasada spójności i zasada niezmienności organizacyjnej. Mając już zbiór fundamentalnych praw fizycznych i psychofizycznych, będziemy mogli w pewnym sensie zrozumieć podstawową strukturę wszechświata.

To trudna sprawa i nie osiągniemy tego w najbliższym czasie. Możemy jednak przynajmniej zmierzać w tym kierunku. Zasady organizacyjnej niezmienności i spójności strukturalnej już nakładają silne ograniczenia na formę teorii fundamentalnej, a nie ma ogromnej liczby potencjalnych podstawowych konstrukcji, które mogą być jej fundamentalnymi składnikami. W tym rozdziale przedstawiam kilka pomysłów, zmierzających w stronę teorii fundamentalnej. Nie prezentuję pełnoprawnej teorii z wy-

czerpującym zbiorem praw podstawowych, ale proponuję pewne konstrukty wchodzące w skład tych praw oraz nakreśliam ogólny możliwy kształt owych praw. Można byłoby to uznać za *prototeorię*: szkielec, na którym można zbudować teorię.

Idee w tym rozdziale są znacznie bardziej szkicowe i spekulatywne niż w pozostałej części książki, stawiają też więcej pytań, niż udzielają odpowiedzi. W ich przypadku prawdopodobieństwo, że są całkowicie błędne, jest największe. Przedstawiając te luźne pomysły, nie zamierzam kreślić ram, które wytrzymają dokładną analizę filozoficzną; są one prezentowane raczej po to, by otwarcie wyłożyć pewne pomysły. Musimy zacząć myśleć o fundamentalnych teoriach świadomości, a może tutaj znajdzie się coś pożytecznego, co pozwoli nam ruszyć do przodu.

## 2 aspekty informacji

Podstawowym pojęciem, jakim będę zajmować się w tym rozdziale, jest pojęcie *informacji*. We współczesnej przestrzeni idei funkcjonuje wiele różnych pojęć informacji, mówiąc zatem o tej ostatniej, trzeba najpierw wyjaśnić, co się ma na myśli. Pojęcie informacji, którym się zajmuję, ma wiele wspólnego z pojęciem omawianym przez Shannona (1948). Tu przedstawiam tę ideę w postaci rozwiniętej i przystosowanej.

Rozwinięcie pozostaje stosunkowo nieformalne, zawiera bowiem tylko tyle formalizmu, [ile trzeba,] aby uchwycić najbardziej podstawowe aspekty pojęcia, które będą istotne. W tym podrozdziale jest trochę technikałów, ale późniejsze podrozdziały są prostsze.

Shannon nie zajmował się semantycznym pojęciem informacji, zgodnie z którym informacje są zawsze informacjami o czymś. Skupił się raczej na formalnym lub składniowym pojęciu informacji, gdzie kluczem jest pojęcie stanu wybranego z repertuaru możliwości. Najbardziej podstawowym ro-

dzajem informacji jest *bit*, który reprezentuje wybór między dwoma możliwościami: mówi się, że jeden bit (0 lub 1) wybrany z dwustanowej przestrzeni przynosi informację. W przypadku bardziej złożonym „komunikat” taki jak „0110010101” wybrany z przestrzeni możliwych komunikatów binarnych przynosi informację w podobny sposób. W ujęciu Shannona *interpretacja* tych stanów nie jest ważna; liczy się *specyficzność* stanu w przestrzeni różnych możliwości.

Możemy sformalizować ten pomysł, odwołując się do pojęcia *przestrzeni informacyjnej*. Przestrzeń informacyjna jest abstrakcyjną przestrzenią złożoną z wielu stanów, które nazywam *stanami informacyjnymi*, oraz z podstawowej struktury *relacji różnicy* między tymi stanami. Najprostszą niebanalną przestrzenią informacyjną jest przestrzeń składająca się z dwóch stanów z pierwotną różnicą między nimi. Możemy ujmować te stany jako dwa „bity”, 0 i 1. Fakt, że oba różnią się od siebie, wyczerpująco opisuje ich naturę. Oznacza to, że przestrzeń informacyjna jest w pełni scharakteryzowana przez jej strukturę różnic.

Inne przestrzenie informacyjne są bardziej złożone. Złożoność może pojawić się z dwóch powodów: albo ze względu na dopuszczenie bardziej skomplikowanej struktury różnic między stanami, albo dopuszczenie, by same te stany miały strukturę wewnętrzną. Chcąc dać przykład pierwszej metody, możemy przenieść się do czterostanowej przestrzeni ze stanami 0, 1, 2 i 3. Przykładem drugiej metody jest ustrukturyzowana przestrzeń ze stanami takimi jak „110010101”. Oczywiście te dwie metody można połączyć, co daje podwójnie złożone stany w przestrzeni komunikatów, takie jak „233102032”. Dalej omówię dokładniej te dwa rodzaje złożoności.

Zacznijmy od pierwszego rodzaju: istnieją oczywiście przestrzenie trójstanowe, czterostanowe i tak dalej, których struktura różnic jest naturalnym rozszerzeniem struktury przestrzeni dwustanowej. Na przykład, element A, B, C lub D wybrany z przestrzeni czteroelementowej przynosi informację w taki sam sposób, jak bit. Oczywiście, natura etykiet „A”, „B” itd. nie jest tutaj istotna; ponownie istotą przestrzeni jest jej struktura.

Co ważniejsze, istnieją *ciągłe* przestrzenie informacyjne, których stany leżą na kontinuum analogicznym do kontinuum liczb rzeczywistych między 0 a 1. Taka przestrzeń ma nieskończenie wiele stanów. Jej struktura jest znacznie bardziej skomplikowana niż w poprzednich przypadkach: bezpośrednio odpowiada topologii kontinuum, z pewnymi stanami leżącymi pomiędzy innymi stanami, pewnymi stanami położonymi bliżej siebie niż innych stanów i tak dalej. Tak jak poprzednio jednak, możemy postrzegać jeden punkt wybrany z kontinuum jako przenoszący informację.

Można również mieć przestrzeń informacyjną, której strukturą jest dwuwymiarowe kontinuum lub wielowymiarowe kontinuum, analogicznie do struktury regionu przestrzeni  $n$ -wymiarowej. Na przykład jeden punkt wybrany z regionu przestrzeni trójwymiarowej będzie przenosić informację. W przypadku najbardziej ogólnym struktura może być podana przez strukturę dowolnej przestrzeni topologicznej, co w rezultacie daje zbiór z relacjami „bliskości” i „sąsiedztwa”. Te szczegóły nie będą jednak zbyt istotne dla reszty wywodu, gdzie zajmują się intuicyjnie znanymi strukturami, takimi jak kontinuum.

Drugi rodzaj złożoności dotyczy stanów mających *strukturę wewnętrzną*. Składają się one z wielu bardziej podstawowych stanów, które nazwę *elementami*. Przykładem jest przestrzeń stanów dziesięciobitowych, analogiczna do „komunikatów” w rodzaju „1001101000”. Każdy stan składa się tutaj z dziesięciu elementów, a każdy element można uznać za podpadający pod swoją własną dwustanową *podprzestrzeń* odpowiadającą oryginalnej przestrzeni dwustanowej. Tę przestrzeń informacyjną możemy postrzegać jako pewnego rodzaju iloczyn dziesięciu podprzestrzeni, z których każda jest przestrzenią informacyjną samą w sobie.

Mogą również istnieć ciekawsze struktury wewnętrzne. Stan informacyjny może mieć na przykład ciągłą strukturę wewnętrzną, tak że jest to rodzaj ciągłego odpowiednika dziesięcioelementowej struktury omówionej wcześniej. Taki stan będzie miał nieskończenie wiele elementów, z których każdy należy do własnej podprzestrzeni. Możemy ujmować odpowiednią

przebieg informacyjną jako zbliżoną do przestrzeni funkcji nad kontinuum (z każdą wartością należącą do podprzestrzeni) lub nad bardziej złożoną przestrzenią ciągłą.

Może się też zdarzyć, że *podprzestrzenie* są złożone w pierwszy wspomniany wyżej sposób: na przykład elementy w każdej podprzestrzeni mogą odpowiadać kontinuum. Jest więc tutaj miejsce na dwa jednocześnie poziomy złożoności. Na przykład każdy stan może składać się z ciągłej struktury elementów, z których każdy może przyjmować wartości w ramach ciągłej podprzestrzeni. Stan informacyjny w tej przestrzeni można postrzegać jako falę lub pewną inną funkcję z ciągłą dziedziną i przeciwdziedziną: jest to ciągle odpowiednik dyskretnych „komunikatów” opisanych wcześniej.

W przypadku najbardziej ogólnym przestrzeń informacyjna będzie miała dwa rodzaje struktury: każdy złożony stan może mieć strukturę wewnętrzną, a każdy element w tym stanie będzie należał do podprzestrzeni z własną topologiczną strukturą różnic. Pierwszą z nich możemy nazwać strukturą *kombinatoryczną* przestrzeni, a drugą – strukturą *relacyjną* podprzestrzeni. Najczęściej każda podprzestrzeń będzie miała taką samą strukturę relacyjną, możemy więc mówić po prostu o strukturze relacyjnej samej przestrzeni.

*Całościowa* struktura przestrzeni jest dana łącznie przez te struktury kombinatoryczne i relacyjne. Często ograniczam uwagę do przestrzeni informacyjnych zawierających tylko strukturę relacyjną, a nie kombinatoryczną – czyli do przypadku, w którym istnieje tylko jeden element w stanie informacyjnym – gdyż wywody są wtedy znacznie prostsze.

Ramy te nie zawierają niczego podobnego do pojęcia treści semantycznej, więc omawiany tutaj rodzaj informacji jest co najwyżej pośrednio związany z semantyczną odmianą informacji omawianą przez filozofów takich jak Dretske (1981) oraz Barwise i Perry (1983). Możliwe jest rozszerzenie istniejących ram teoretycznych w stronę elementu semantycznego przez przyporządkowanie jakiejś treści semantycznej każdemu stanowi in-

formacyjnemu, ale w obecnej formie rami te są niezależne od rozważań semantycznych.

Ta formalizacja wyraża ideę Shannona, że informacja polega przede wszystkim na stanie wybranym z wielu możliwości (w strukturze relacyjnej przestrzeni), a także oddaje ideę, że informacje złożone można zbudować z prostych (w kombinatorycznej strukturze przestrzeni). Dla Shannona jeden bit może stanowić informację, podobnie jak długi „komunikat” w rodzaju „10011010”. Shannon rozważa również przypadek, gdy informacje należą do przestrzeni ciągłej lub przestrzeni funkcji nad ciągłą dziedziną. W każdym przypadku zasadnicze znaczenie ma wybór jednego elementu z przestrzeni kontrastujących możliwości.

Ujęcie samego Shannona dotyczy często *ilości informacji zawartych w stanie informacyjnym, co stanowi miarę specyficzności stanu* w przestrzeni informacyjnej. Stan w dwustanowej przestrzeni informacyjnej przenosi jeden bit informacji; stan w czterostanowej przestrzeni przenosi dwa bity; stan w przestrzeni  $n$ -elementowej przenosi  $\log_2 n$  bitów. Kiedy przestrzeń jest kombinacją podprzestrzeni, stan przenosi ilość informacji równą sumie ilości przesyłanych przez jego elementy: tak więc dziesięciocyfrowy „komunikat” binarny przenosi dziesięć bitów informacji. Ten zabieg stosuje się do przestrzeni dyskretnych; w przestrzeni ciągłej ilość informacji musi być określona bardziej subtelnie. Nie będę się tutaj zbytnio zajmował ilością informacji, ale raczej samymi stanami informacyjnymi, które można uważać za pozostające w takiej samej relacji z ilością informacji, w jakiej materia pozostaje z masą.

### informacje realizowane fizycznie

Zdefiniowane przeze mnie przestrzenie informacyjne są przestrzeniami *abstrakcyjnymi*, a stany informacyjne są stanami abstrakcyjnymi. Nie należą do konkretnego świata fizycznego ani fenomenalnego. Możemy jednak znaleźć informacje zarówno w świecie fizycznym, jak i fenomenalnym, jeśli

spojrzymy na sprawy we właściwy sposób. W tym celu musimy omówić różne sposoby *realizacji* przestrzeni i stanów informacyjnych w świecie. Omówię po kolei realizację fizyczną i fenomenalną.

Wydaje się intuicyjnie oczywiste, że przestrzenie i stany informacyjne są realizowane w całym świecie fizycznym. *Na przykład mój włącznik światła można postrzegać jako realizujący dwustanową przestrzeń informacyjną, a jego stany „w górę” i „w dół” jako realizacje obu stanów. Dysk kompaktowy zaś można postrzegać jako realizujący kombinatoryczny stan informacyjny, polegający na złożonej strukturze bitów.* Podobnie można postrzegać informacje realizowane w termostacie, książce lub linii telefonicznej. Jak rozumieć te intuicje?

Naturalnym sposobem powiązania układów fizycznych i stanów informacyjnych jest postrzeganie fizycznie zrealizowanych informacji w myśl sloganu ukutego przez Batesona (1972): *informacja to różnica, która czyni różnicę*. Podczas gdy mój włącznik światła może przyjmować nieskończoną liczbę pozycji w ciągłym zakresie, większość tych zmian nie ma znaczenia dla oświetlenia. Światło będzie włączone bez względu na to, czy przełącznik jest całkowicie podniesiony, czy opuszczony o jedną czwartą w dół. Z drugiej strony, gdy jest w pozycji bardziej niż około jedna trzecia w dół, światło zgaśnie. Jeśli chodzi o oświetlenie, istnieją tylko dwa istotne stany włącznika, które można nazwać „góra” i „dół”. Różnica między nimi jest jedyną różnicą, która czyni różnicę dla oświetlenia. Widzimy więc, że włącznik realizuje dwustanową przestrzeń informacyjną, z  *pewnymi swoimi stanami fizycznymi odpowiadającymi jednemu stanowi informacyjnemu, a innymi – drugiemu.*

Ogólnie rzecz biorąc, przestrzeń informacyjna związana z obiektem fizycznym będą zawsze zdefiniowana w odniesieniu do *ścieżki przyczynowej* (w tym przypadku ścieżki od włącznika światła do oświetlenia) i przestrzeni możliwych *efektów* na końcu ścieżki (w tym przypadku stan włączenia lub wyłączenia oświetlenia). Stany fizyczne będą odpowiadać stanom informacyjnym zgodnie z ich efektami na ścieżce przyczynowej. Kiedy dwa stany

fizyczne mają ten sam efekt na ścieżce – tak jak dwie pozycje włącznika światła, które tak samo prowadzą do włączenia oświetlenia – będą odpowiadać temu samemu stanowi informacyjnemu. Jeśli w ten sposób podzielimy stany fizyczne, dojdziemy do podstawowego zbioru różnic fizycznych, które czynią różnicę, zbioru, składającego się na fizyczną realizację przestrzeni informacyjnej.

**Struktura przestrzeni informacyjnej będzie odpowiadać bezpośrednio strukturze przestrzeni efektów, która sama będzie przestrzenią dyskretną lub ciągłą. W przypadku oświetlenia istnieją na przykład dwa istotne efekty na ścieżce przyczynowej: oświetlenie może być włączone lub wyłączone. Włącznik można więc postrzegać jako realizację dwustanowej przestrzeni informacyjnej.**

Ciągłe przestrzenie informacyjne można traktować w podobny sposób. Jeśli włącznik jest wyposażony w ściemniacz, to obrócenie pokrętki na różne pozycje skutkuje różnym natężeniem światła w ciągłym zakresie. (W praktyce zakres może być dyskretny, ale tu idealizuję tę sytuację). Wpływ na natężenie światła definiuje ciągłą przestrzeń informacyjną zrealizowaną w moim włączniku światła. Stany fizyczne włącznika, które wywołują takie samo natężenie światła (być może stany w obszarach, w których pokrętko jest niewrażliwe lub stany, które różnią się nieistotnymi parametrami, takimi jak kolor włącznika), będą przyporządkowane temu samemu stanowi informacyjnemu. Przestrzeń stanów informacyjnych ma topologiczną strukturę kontinuum, ze strukturą różnic między stanami odpowiadającą strukturze różnic efektów wpływających na natężenie światła.

Informacje zrealizowane w dysku kompaktowym można także zanalizować w ten sposób. Dysk ma nieskończenie wiele możliwych stanów fizycznych, ale gdy bierzemy pod uwagę jego wpływ na odtwarzacz kompaktowy, dysk realizuje tylko skończoną liczbę możliwych stanów informacyjnych. Wiele zmian na dysku – zmiany mikroskopowe poniżej poziomu rozdzielczości czytnika optycznego, ryska na dysku lub duży znak na odwrocie – nie powodują różnic w funkcjonowaniu systemu. Jedyne różnice istotne

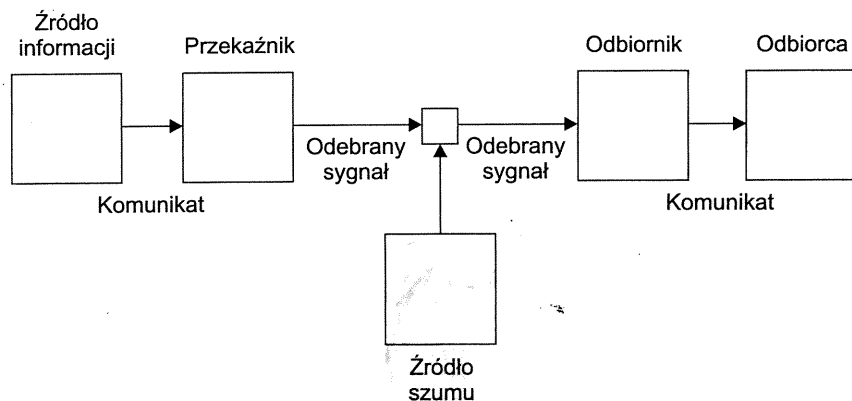
dla stanu informacyjnego dysku to takie, które odzwierciedlają się w wyjściu czytnika optycznego. Są to różnice w obecności wgłębień i pól na dysku, odpowiadające temu, co uważamy za „bity”. Dowolny stan dysku będzie miał przypisany stan informacyjny w dużej przestrzeni informacyjnej. **Fizyczne stany różnych tłoczeń tego samego nagrania będą miały przypisany ten sam stan informacyjny, jeśli wszystko pójdzie dobrze.** Tłoczenia różnych nagrań, a nawet niedoskonałe tłoczenia tego samego nagrania, będą miały przypisane różne stany informacyjne ze względu na ich różne efekty.

Jest to przypadek, w którym fizycznie zrealizowana przestrzeń informacyjna ma strukturę kombinatoryczną. Każdy „bit” na dysku kompaktowym ma niezależny efekt wpływający na odtwarzacz dysków kompaktowych, więc każde miejsce na dysku można postrzegać jako realizację własnej dwustanowej podprzestrzeni. Łącząc wszystkie te niezależne efekty, odkrywamy strukturę kombinatoryczną w przestrzeni łącznych efektów płyty kompaktowej, a więc odnajdujemy tę samą strukturę kombinatoryczną w przestrzeni informacyjnej, którą realizuje płyta. Tę przestrzeń informacyjną można postrzegać jako iloczyn dużego zbioru dwustanowych podprzestrzeni, po jednej dla każdego wgłębienia i pola na dysku.

Zauważmy, że w tym ujęciu fizycznie zrealizowana informacja jest informacją tylko o tyle, o ile można ją przetwarzać. Jak ujął to Mackay (1969), „[I]nformacja jest tym, co informacja powoduje”. Wpisuje się to w koncepcję informacji Shannona. „Ilość informacji” u tego ostatniego stanowi miarę **specyficzności stanu w przestrzeni stanów, które mogą być przekazane – to znaczy, które mogą odgrywać różne role na różnych ścieżkach przyczynowych (zwanym przez Shannona kanałem komunikacyjnym).** Dla Shannona informacja jest zawsze stanem przekazywalnym, a **wielkość przestrzeni informacyjnej jest w istocie niejawnie zdefiniowana przez funkcję nadajnika.** Informacja jest różnicą, która może powodować różnicę w transmisji.

Jasno widać to na standardowym diagramie Shannona (rys. 8.1) i związanych z nim objaśnieniach. Źródło informacji to zbiór „komunikatów” (stanów informacyjnych), w którego przypadku zakłada się, że osobne komuni-

katy kodowane są przez nadajnik na osobne sygnały, a osobno przesyłane sygnały odpowiadają różnym komunikatom.



**Rysunek 8.1** Diagram kanału informacyjnego u Shannona (diagram 1 za Claude E. Shannon i Warren Weaver, *Mathematical Theory of Communication*, 1963. Copyright 1949 by the Board of Trustees of the University of Illinois Press. Użyty za zgodą University of Illinois Press.).

W istocie własność ta *definiuje*, kiedy stany liczą się jako osobne komunikaty. Jeśli dwa różne stany fizyczne układu przekształca się na ten sam sygnał, to realizują one ten sam komunikat. Zniekształcenie i utrata informacji mogą pojawić się na dalszym etapie procesu, jak wskazuje prawa strona diagramu, ale konstytutywną cechą kanału informacyjnego jest to, że osobne stany informacyjne wytwarzają różne skutki przez nadajnik. Wszystko to jest zawarte *implicite*, a nie *explicite* w ujęciu Shannona, gdzie nie ma bezpośredniego ujęcia relacji między stanami fizycznymi a stanami informacyjnymi. Jednak przy bliższych oględzinach oczywiste staje się, że gdy stany informacyjne są indywidualizowane, w grę wchodzi zasada przekazywalności.

Nie będę próbował podawać dokładnych kryteriów realizacji przestrzeni informacyjnej w układzie fizycznym. Pozostawię sprawę na nieformalnym

poziomie zasady „różnicy, która czyni różnicę”. Tę nieformalną ideę można różnie formalizować, a pewne formalizacje mocniej ograniczają realizację niż inne. Zbyt wcześnie byłoby rozstrzygać na korzyść którejkolwiek z nich, w szczególności w tym momencie. Pozostawiając kwestię w nieformalnym ujęciu, dopuszczamy pewne pole manewru, gdy chodzi o szczegóły, które można dopowiedzieć, kiedy będziemy mieli lepsze pojęcie o tym, co jest właściwe dla określonego zastosowania. W przypadku teorii świadomości wypełnienie tych szczegółów treścią w odpowiedni sposób należy do procesu konkretyzacji teorii.

### informacje realizowane fenomenalnie

Realizacja fizyczna jest najbardziej powszechnym sposobem ujmowania informacji osadzonych w świecie, ale informacje ujawniają się nie tylko w taki sposób.

Możemy znaleźć informacje realizowane w naszej *fenomenologii*. Stany przeżyć w sposób naturalny należą bezpośrednio do przestrzeni informacyjnych. Istnieją naturalne wzorce podobieństwa i różnic między stanami fenomenalnymi, tworzące strukturę różnic przestrzeni informacyjnej. Widzimy więc, że stany fenomenalne realizują stany informacyjne w tych przestrzeniach.

Na przykład przestrzeń prostych przeżyć barw ma trójwymiarową strukturę relacyjną, którą już omawialiśmy. Abstrahując wzory podobieństwa i różnic między tymi przeżyciami, otrzymujemy abstrakcyjną przestrzeń informacyjną z trójwymiarową strukturą relacyjną, którą realizuje przestrzeń fenomenalna. Dowolne proste przeżycie barwy odpowiada konkretnemu miejscu w tej przestrzeni. Konkretnie przeżycie czerwieni jest jednym fenomenalnie zrealizowanym stanem informacyjnym, a konkretne przeżycie zieleni – drugim.

Bardziej skomplikowane przeżycia, takie jak przeżycie całego pola widzenia, należą do przestrzeni informacyjnych o złożonej strukturze

świadomość i informacja: kilka spekulacji

kombinatorycznej. Kiedy na przykład patrzę na obraz, moje przeżycie należy do przestrzeni ze strukturą kombinatoryczną (co najmniej) dwuwymiarowego kontinuum, gdzie każdy element w tym kontinuum ma (co najmniej) trójwymiarową strukturę relacyjną prostej przestrzeni barw. Struktura plam barwnych w polu widzenia nie jest *tak* odmienna od struktury cyfr binarnych w dziesięciocyfrowym komunikacie, chociaż zarówno struktura kombinatoryczna, jak i relacyjna są w tym wypadku znacznie bardziej skomplikowane.

Szukając przestrzeni informacyjnych realizowanych fenomenalnie, nie opieramy się na przyczynowej zasadzie „różnicy, która czyni różnicę”, jaką stosowaliśmy do znajdowania przestrzeni informacyjnych zrealizowanych fizycznie. Opieramy się raczej na wewnętrznych jakościach przeżyć i strukturze istniejącej między nimi – relacjach podobieństwa i różnicy, jakie zachodzą między nimi, a także na ich wewnętrznej strukturze kombinatorycznej. Każde przeżycie będzie wchodzić w naturalne relacje podobieństwa i różnicy z innymi przeżyciami, dlatego zawsze będziemy w stanie znaleźć przestrzenie informacyjne, do których należą przeżycia.

### zasada podwójnego aspektu

Takie ujęcie informacji ujawnia istotne ogniwo łączące świat fizyczny i fenomenalny: **ilekroć znajdujemy przestrzeń informacyjną zrealizowaną fenomenalnie, znajdujemy tę samą przestrzeń informacyjną zrealizowaną fizycznie.** A kiedy przeżycie realizuje stan informacyjny, ten sam stan informacyjny jest realizowany w fizycznym substracie przeżycia.

Weźmy proste przeżycie barwy, realizujące stan informacyjny w trójwymiarowej przestrzeni informacyjnej. Możemy znaleźć tę samą przestrzeń zrealizowaną w procesach mózgowych leżących u podstaw przeżycia: jest to trójwymiarowa przestrzeń neuronowo kodowanych reprezentacji w korze wzrokowej. Elementy tej trójwymiarowej przestrzeni odpowiadają bezpośrednio elementom fenomenalnej przestrzeni informacyjnej.

Nie wiemy dokładnie, jak te stany są kodowane, a więc nie wiemy dokładnie, jak przestrzeń informacyjna jest zrealizowana fizycznie. Wiemy jednak, że musi być zrealizowana, gdyż w późniejszych procesach ujawniają się wszelkie systematyczne efekty realizacji informacyjnej. Nasze raporty mogą się na przykład systematycznie różnić co do lokalizacji w przestrzeni barw, np. wtedy, gdy oceniamy barwy jako względnie „ciemniejsze” i „jaśniejsze”; możemy dopasować przedmioty do innych przedmiotów na podstawie ich podobieństw i różnic pod względem barwy. Wiemy więc, że muszą istnieć istotne różnice w korze wzrokowej, które są *przekazywalne* do innych obszarów mózgu, tworząc trójwymiarową przestrzeń możliwych wpływów. Stany leżące u podstaw dowolnych dwóch nieodróżnialnych przeżyć będą miały takie same istotne efekty, nawet jeśli wiążą się z nieco innymi szczegółami fizycznymi – mamy tu analogię z niewielkimi różnicami w stanie włącznika światła – a stany leżące u podstaw dowolnych dwóch *podobnych* przeżyć będą miały podobne efekty. Możemy więc postrzegać korę wzrokową jako realizującą stany informacyjne w przestrzeni trójwymiarowej.

To samo odnosi się do bardziej skomplikowanych przeżyć, takich jak przeżycia całego pola widzenia. Są one realizowane w kombinatorycznej przestrzeni informacyjnej, a sama ta przestrzeń musi być fizycznie zrealizowana w leżących u ich podstaw procesach mózgowych. Wiemy, że dla każdej lokalizacji w polu, różne proste przeżycia odpowiadają różnicom w różnych późniejszych efektach, jeśli te późniejsze różnice można rozdzielić w zależności od lokalizacji w polu. Na przykład możemy odpowiedzieć osobno na konkretne pytania o kolor w danej lokalizacji; ta osobna przestrzeń efektów dla każdej lokalizacji daje oddzielną podprzestrzeń dla każdej lokalizacji. Tak więc gdzieś w korze wzrokowej musi następować kodowanie kombinatorycznego stanu informacyjnego, aby wszystkie istotne różnice mogły być przekazywane późniejszym procesom. **Przestrzeń odpowiednich możliwych stanów jest tutaj izomorficzna z przestrzenią możliwych przeżyć;** możemy więc zobaczyć, że ten sam stan informacyjny jest zrealizowany zarówno fizycznie, jak i fenomenalnie.

Informacje te nie muszą być zakodowane *lokalnie*, na przykład w małej strukturze sąsiadujących neuronów. Jest całkiem możliwe, że informacje fizyczne zrealizowane są w sposób holistyczny, tak jak na przykład w pewnych holograficznych postaciach przechowywania informacji. Istotnym różnicom w stanach kory wzrokowej mogą odpowiadać różnice rozproszone w różnych częściach kory mózgowej. Dopóki jednak te różnice są przekazywane i mają istotne efekty, dopóty informacje te i tak będą realizowane.

Naturalne jest przypuszczenie, że to podwójne życie przestrzeni informacyjnych odpowiada dwoistości na głębokim poziomie. Można by nawet sugerować, że ta podwójna realizacja jest kluczem do fundamentalnego związku między procesami fizycznymi a przeżyciem świadomym. Potrzebujemy *jakiegoś* rodzaju konstrukcji, aby je powiązać, a informacja wydaje się konstruktem co najmniej równie dobrym, jak inne. Możliwe, że zasady dotyczące podwójnej realizacji informacji mogłyby zostać skonkretyzowane jako system podstawowych praw łączących dziedzinę fizyczną i fenomenalną.

Możemy to wyrazić, proponując jako podstawową zasadę to, że informacja (w rzeczywistym świecie) ma dwa aspekty, fizyczny i fenomenalny. **Wszędzie tam, gdzie jest stan fenomenalny, realizuje on stan informacyjny, który jest również zrealizowany w systemie poznawczym mózgu.** Przynajmniej zaś dla niektórych fizycznie zrealizowanych przestrzeni informacyjnych, ilekroć stan informacyjny w tej przestrzeni jest realizowany fizycznie, jest także realizowany fenomenalnie<sup>1</sup>.

Zasada ta sama z siebie nie stanowi bynajmniej pełnej teorii psychofizycznej. Stanowi raczej swego rodzaju szablon dla teorii psychofizycznej, zapewniając podstawowe ramy, w których można ujmować szczegółowe prawa. Konkretyzując zasadę do postaci teorii, należy odpowiedzieć na

<sup>1</sup> Powiązania między świadomością a informacją sugerowali także Bohm (1980), Sayre (1976) i Velmans (1991), jednak szczegóły ich propozycji są zupełnie odmienne od moich. Idea „monizmu neutralnego” informacji u Sayre’a jest mimo to bardzo sugestywna (dziękuję Steve’owi Horstowi za tę wskazówkę). Coś w rodzaju zasady podwójnego aspektu omawia Lockwood 1989, rozdz. 11, chociaż nie wystawia tego w kategoriach informacji.

różnego rodzaju pytania. Na przykład, do *jakich* fizycznie zrealizowanych przestrzeni informacyjnych stosuje się podstawową zasadę? Omówię to pytanie bardziej szczegółowo w podrozdziale 4, ale na razie pozostawiam je otwarte. Innego rodzaju niedookreślenie wynika z luźnej definicji fizycznie zrealizowanej informacji: aby mieć w pełni skonkretyzowaną teorię psychofizyczną, musimy dokładnie wiedzieć, co to znaczy, że przestrzeń informacyjna jest fizycznie zrealizowana. Wszystko to należy jednak do procesu opracowywania teorii.

Inne ważne pytania dotyczą *ontologii* tego stanowiska. Jak poważnie powinniśmy potraktować wyrażenie „podwójny aspekt”? W jakim stopniu w tych ramach *reifikuje* się informację, czyli traktuje ją jako realną? Czy twierdzi się, że to, co fizyczne, i to, co fenomenalne, (albo oba), są ontologicznie zależne od tego, co informacyjne? Wszystkie te pytania pozostaną na razie otwarte. Później w tym rozdziale rozważę różne możliwości interpretacji ontologicznych. Niektóre z tych interpretacji uznają po prostu informację za konstrukt przydatny do charakterystyki praw psychofizycznych; inne przyznają jej bardziej zasadniczą rolę w ontologii. Podobnie niektóre interpretacje traktują ideę „podwójnego aspektu” poważniej niż inne.

Na razie abstrahuję od tych zagadnień metafizycznych. Moją zasadę należy po prostu traktować jako prawo łączące dziedzinę fizyczną i fenomenalną z implikacjami ontologicznymi, które nie różnią się szczególnie od tych, z jakimi mamy do czynienia w przypadku wcześniej rozpatrywanych praw. Wiemy już, że przeżycie powstaje z tego, co fizyczne, na mocy pewnych praw, które mają zastosowanie do pewnych cech fizycznych świata. Najważniejszą sugestią jest to, że podstawowym poziomem, na którym prawa stosują się do świata fizycznego, jest poziom fizycznie zrealizowanej informacji.

Oczywiście, informacja nie może być *pierwotną* cechą świata fizycznego w taki sposób, w jaki masa i ładunek są pierwotne, ale nie musi być pierwotna, by odgrywać rolę w fundamentalnych prawach psychofizycz-

nych. Mamy już wszystkie potrzebne fundamentalne własności w postaci podstawowych własności fizycznych i fenomenalnych. Potrzebujemy teraz konstrukt łączącego te dziedziny. Informacja wydaje się prostym i jasnym konstruktem, który doskonale nadaje się na tego typu łącznik i który może rzeczywiście przynieść w efekcie zbiór prostych i wyczerpujących zarazem praw. Jeśli taki zbiór można uzyskać, to możemy mieć naprawdę fundamentalną teorię świadomości.

Możliwe jednak, że samą informację można postrzegać jako fundamentalną. Na przykład idea, że fizyka zajmuje się w ostatecznym rozrachunku informacją, była już szkicowana przez niektórych fizyków. Gdyby tę ideę można było rozwinąć, mogłoby się okazać, że w jakiś sposób to, co fizyczne, jest pochodną tego, co informacyjne, a ontologia tego stanowiska mogłaby zostać opracowana bardzo elegancko. Omówię kilka idei utrzymanych w tym duchu w dalszej części rozdziału.

### 3 argumenty na rzecz teorii

Nie mam żadnych powalających argumentów, które dowodziłyby, że informacja jest kluczem do powiązania procesów fizycznych i przeżyć świadomych, ale ideę tę można wesprzeć w pewien pośredni sposób. Pisałem już o pomocniczym względzie pierwszego rodzaju: obserwacji, że te same przestrzenie informacyjne są realizowane fizycznie i fenomenalnie. Niżej wspomnę o innym ważnym źródle wsparcia i dwóch innych źródłach mających mniejsze znaczenie.

Dwa źródła wsparcia o mniejszym znaczeniu to fakt, że ujmowanie informacji jako mającej podwójny aspekt jest zgodne z zasadami psychofizycznymi opracowanymi wcześniej: w szczególności z zasadą spójności strukturalnej i zasadą niezmienności organizacyjnej. Zasady te są silnymi ograniczeniami i nie jest oczywiste, jak teoria fundamentalna może je

uwzględnić, a więc świadczy to na korzyść poglądu informacyjnego, który jest zgodny z oboma.

Zgodność z zasadą spójności strukturalnej jest szczególnie dobrze widoczna: pod pewnymi względami pogląd informacyjny jest idealnie dostosowany do tego, by spełniać to ograniczenie. Struktura przeżycia jest dokładnie strukturą fenomenalnie zrealizowanej przestrzeni informacyjnej, a struktura przytomności jest dokładnie strukturą fizycznie zrealizowanej przestrzeni informacyjnej. Aby dostrzec słuszność pierwszej obserwacji, zauważmy, że to, co nazwałem niejawną strukturą przeżycia, odpowiada relacyjnej strukturze przestrzeni informacyjnej, a to, co nazwałem jawną strukturą przeżycia, odpowiada kombinatorycznej strukturze przestrzeni. Aby przekonać się o słuszności drugiej obserwacji, zauważmy, że różne szczegóły struktury przytomności są z definicji różnicami, które czynią różnicę w dalszym przetwarzaniu, ponieważ są bezpośrednio dostępne dla globalnej kontroli, a więc stanowią fizyczną realizację przestrzeni informacyjnej. Skoro obie są faktycznie realizacjami tej samej przestrzeni informacyjnej, wynika stąd zasada strukturalnej spójności.

Powinienem zaznaczyć, że zasada podwójnego aspektu sama z siebie nie zapewnia, iż struktura przytomności będzie rzutowana na przeżycie. Aby się co do tego upewnić, musimy pokazać, że dana fizyczna przestrzeń informacyjna jest jedną z tych, do których stosuje się zasadę podwójnego aspektu. W tym celu będziemy potrzebować dokładniejszej wersji zasady, która zawęży wchodzące w grę przestrzenie informacyjne w odpowiedni sposób, aby przynajmniej zawierały informacje, które są udostępniane dla globalnej kontroli w znanych przypadkach. W obecnej formie zasada podwójnego aspektu jeszcze nie przewiduje pełnej zasady strukturalnej spójności, ale przynajmniej jest z nią zgodna.

Nie jest również trudno zauważyć, że zasada podwójnego aspektu jest zgodna z zasadą niezmienności organizacyjnej. Aby się o tym przekonać, zauważmy, że gdy system realizuje przestrzeń informacyjną, czyni to na mocy swojej organizacji funkcjonalnej. Każdy inny system, który jest funk-

świadomość i informacja: kilka spekulacji

cjonalnie izomorficzny z wystarczająco wysoką dokładnością, będzie miał dokładnie ten sam wzorec różnic, które czynią różnicę, a więc będzie realizować tę samą przestrzeń informacyjną. Jeśli więc moje przeżycia powstają na mocy przestrzeni informacyjnych zrealizowanych w moim mózgu, to te same przestrzenie informacyjne będą realizowane w odpowiedniku izomorficznym funkcjonalnie i powstaną te same przeżycia, jak przewiduje to zasada niezmienności.

Fundamentalna teoria świadomości będzie musiała przywołać cechy fizyczne, które są zarówno organizacyjnie niezmiennie, jak i wystarczająco proste, by odgrywać rolę w prawach fundamentalnych. Większość cech organizacyjnie niezmiennych nie jest bardzo prosta, a najprostsze cechy nie są organizacyjnie niezmiennie. Fizycznie zrealizowane informacje mogą być najbardziej naturalną cechą, która spełnia oba kryteria. Fakt, że spełniają oba, świadczy na korzyść informacyjnego podejścia do praw psychofizycznych.

### wyjaśnianie sądów fenomenalnych

Wcześniej widzieliśmy, że chociaż świadomość nie może być wyjaśniona redukcjonnie, sądy fenomenalne – sądy o postaci „Jestem świadomy”, „Czyż świadomość nie jest dziwna” itd. – mogą być tak wyjaśnione, przynajmniej co do zasady. To obciążało w pewien sposób nieredukcyjną teorię świadomości, chociaż nie wydawało się zgubne. Jest sprzeczne z intuicją, że sądy te można wyjaśnić bez odwoływania się do samej świadomości, ale możemy nauczyć się z tym żyć. Zarazem jednak możemy mieć nadzieję, że wyjaśnienie sądów fenomenalnych będzie związane w jakiś głębszy sposób z wyjaśnieniem samej świadomości. Wydaje się nieuzasadnione i przypadkowe, że te dwa wyjaśnienia pozostają całkowicie niezależne.

Możemy to ująć jako pewnego rodzaju wymóg *spójności eksplanacyjnej* wobec teorii świadomości. Kompletna teoria umysłu musi zapewnić zarówno (nieredukcyjne) ujęcie świadomości, jak i (redukcyjne) ujęcie tego, dlaczego

osądzamy, że mamy świadomość; rozsądne jest oczekiwanie, że te dwa ujęcia będą *spójne* ze sobą. W szczególności można by się spodziewać, że te cechy przetwarzania, które są przede wszystkim odpowiedzialne za wywołanie sądów fenomenalnych, będą również w pierwszym rzędzie odpowiedzialne za samą świadomość. W ten sposób, nawet jeśli sama świadomość nie pojawia się w wyjaśnieniu sądów fenomenalnych, korzenie świadomości będą się w nim pojawiać.

Oczywiście nie można *udowodnić*, że teoria świadomości musi spełnić ten wymóg, ale dowolna teoria świadomości, która go spełni, będzie mieć przewagę nad teoriami, które go nie spełnią. Jeśli teoria pokazuje, w jaki sposób wyjaśnienie sądów fenomenalnych zasadniczo wymaga eksplanacyjnej podstawy świadomości, to łączymy obie [te sfery] w bardziej jednolitym obrazie umysłu i znika poczucie skandalicznego przypadku.

Często sądziłem, że może to być klucz do znalezienia teorii świadomości<sup>2</sup>. Po pierwsze, musimy jak najlepiej zrozumieć, dlaczego tworzone są sądy o świadomości. To może być trudne pytanie, ale nie powinno dotyczyć głębokich metafizycznych tajemnic; w zasadzie należy ono do dziedziny kognitywistyki. Następnie musimy wyabstrahować najważniejsze cechy tego wyjaśnienia i zastanowić się, jak mogą one odgrywać istotną rolę w teorii samej świadomości. Nie ma gwarancji, że doprowadzi to do stworzenia zadowalającej teorii świadomości, ale jest obiecującą strategią.

Poszukiwanie redukcyjnych wyjaśnień naszych sądów o świadomości będzie w każdym razie chyba pouczające i należy do najbardziej opłacalnych zastosowań metod redukcyjnych w pracach nad teorią świadomości. Możemy skoncentrować się na pytaniu, dlaczego system przetwarzania ma tworzyć sądy, że występuje świadomość, a w szczególności, dlaczego ma sądzić, że świadomość jest takim dziwnym zjawiskiem. Powiedziałem

2 Mój niepublikowany artykuł (Chalmers 1990) skupia się na tej strategii rozumienia relacji między świadomością a sądami na temat świadomości i stosuje ją przy opracowywaniu pewnej podstawowej „teorii” świadomości (zawierającej [pojęcia] wzorca i informacji), stanowiącej pierwotną wersję niektórych idei z tego rozdziału. W tym artykule wymóg spójności eksplanacyjnej nazywam „testem spójności”, który musi przejść każda teoria świadomości.

już parę słów w tym duchu w rozdziale V; tu wejść w szczegóły. „Wyjaśnienie”, które podam, jest jedynie przekonującą bajeczką, ale można mieć nadzieję, że uda się ją skonkretyzować przy użyciu badań empirycznych do postaci szczegółowej teorii. Kognitywistyka i neuronauka prawdopodobnie bardzo skorzystają, starając się uporać z tymi zjawiskami.

Zostawmy więc samą świadomość przez moment na boku i skoncentrujmy się na systemie przetwarzania poznawczego z punktu widzenia trzeciej osoby. Podane wyjaśnienie można ująć jako stosujące się do zombiego, jeśli ktoś chce. Dlaczego mielibyśmy się *spodziewać*, że system przetwarzania stworzy tego rodzaju sąd? Jakiego rodzaju proces może leżeć u podstaw sądu, że na przykład występuje doznanie barwy? Aby się nad tym zastanowić, rozważmy, co może się dziać, gdy widzimy barwy.

Nie wchodząc w szczegóły niskiego poziomu, [możemy powiedzieć, że] rzecz przebiega z grubsza w następujący sposób. Szczególna obwiednia widmowa światła dociera do naszych oczu, aktywując różnego rodzaju komórki siatkówki. Trzy odmiany czopków wyabstrahowują informacje w zależności od ilości światła występującej w różnych nakładających się zakresach długości fal. Natychmiast traconych jest wiele rozróżnień występujących w pierwotnej fali światła. Te informacje są przekazywane dalej przez nerw wzrokowy do kory wzrokowej, gdzie dalej następuje ich neuronowe przetwarzanie na informacje odpowiadające wartościom na trzech osiach – są to zapewne osie: czerwono-zielona, żółto-niebieska i achromatyczna. Słabo rozumiemy to, co dzieje się później, ale wydaje się, że informacje odpowiadające lokalizacji danego koloru w tej trójwymiarowej przestrzeni są zachowywane, zanim ostatecznie zostaną uporządkowane w znanych kategoriach „czerwony”, „zielony”, „brązowy” i tak dalej. Do tych etykiet dołączane są kategorie słowne, a wreszcie wydawany jest raport taki jak „Widzę teraz czerwień”.

Przyjmijmy teraz „punkt widzenia” systemu na to, co się dzieje. Jakiego rodzaju sądy będzie on tworzyć? Z pewnością będą to sądy takie jak „tam jest czerwony obiekt”, ale jeśli system jest racjonalny i refleksyjny, możemy również oczekiwać, że będzie w stanie zastanawiać się również nad samym

procesem postrzegania. Możemy zapytać, w jaki sposób postrzeganie „dociera” do systemu.

Zasadniczą cechą jest tutaj to, że gdy system postrzega czerwony obiekt, centralne procesy nie mają bezpośredniego dostępu do samego obiektu ani do procesów fizycznych leżących u podstaw postrzegania. *Wszystko*, do czego procesy te mają dostęp, to sama informacja o barwach, będąca tylko lokalizacją w trójwymiarowej przestrzeni informacyjnej. Jeśli chodzi o językowe raporty dotyczące sytuacji, system nie może relacjonować „Ta plama jest nasycona odbiciami rzędu 500–600 nanometrów”, gdyż cały dostęp do oryginalnych długości fal zostaje utracony. Podobnie nie może tworzyć relacji na temat struktury nerwowej typu „jest teraz częstotliwość czynnościowa 50 Hz”, gdyż nie ma bezpośredniego dostępu do struktur nerwowych. System ma dostęp tylko do lokalizacji w przestrzeni informacyjnej.

W istocie, jeśli chodzi o przetwarzanie centralne, po prostu *znajduje się* on w lokalizacji w tej przestrzeni. System jest w stanie rozróżniać i *wie*, że jest w stanie to robić, ale nie ma pojęcia, jak to robi. Oczekiwaliśmy, że po chwili zacznie nadawać *etykiety* różnym lokalizacjom, w których się znajduje – „czerwony”, „zielony” i tym podobne – i będzie mógł wiedzieć, w jakim jest stanie w danym czasie. Na pytanie jednak, skąd konkretnie wie, nie może powiedzieć nic ponad i poza tym, że „Po prostu wiem, bezpośrednio”. Jeśli ktoś spyta: „Jaka jest różnica między tymi stanami?”, nie ma na to odpowiedzi poza: „Są po prostu różne” lub: „Jest to jeden z *tych*”, lub: „Ten jest *czerwony*, a ten *zielony*”. Naciskany w kwestii znaczenia tych stwierdzeń, system nie ma nic do powiedzenia poza: „Są po prostu różne, jakościowo”. Cóż innego mógłby powiedzieć?

Naturalne jest przypuszczenie, że system, który może bezpośrednio znać swoją lokalizację w przestrzeni informacyjnej, nie mając dostępu do żadnej dalszej wiedzy, po prostu określi stany jako surowo i pierwotnie różne, różniące się „jakością”. Oczywiście, możemy oczekiwać, że te różnice dotrą do systemu w „bezpośredni” sposób: wchodzi on w te stany, które

z kolei są bezpośrednio dostępne przy kierowaniu dalszym przetwarzaniem; w jego wiedzy na temat stanu, w którym jest, nie ma na przykład nic inferencyjnego. Powinniśmy też oczekiwać, że te stany będą w dość istotnym stopniu „niewystawialne”: system nie ma dostępu do dodatkowych istotnych informacji, więc nie może nic powiedzieć o tych stanach poza wskazaniem na podobieństwa i różnice między nimi, a także na różne skojarzenia z nimi związane. Z pewnością nie oczekivalibyśmy, że „jakość” będzie czymś, co można objaśnić w bardziej podstawowych kategoriach.

Można by wysunąć zarzut, że system można byłoby skonstruować tak, aby miał dostęp do informacji w postaci „przeczuć”, w podobny sposób jak osoba cierpiąca na ślepowidzenie. Być może mógłby powiedzieć: „Sąd «czerwień» po prostu przyszedł mi do głowy”, nie twierdząc nic na temat „jakości”. To prawdopodobnie byłaby jednak konstrukcja nieefektywna, gdyż system musiałby czekać na przecucie. A co dzieje się w sytuacji, kiedy (powiedzmy, w czasie gry w tenisa) trzeba reagować na informacje wzrokowe bez tworzenia sądów? Przypuszczalnie system powiedziałby: „Właśnie zacząłem wiedzieć, gdzie była piłka i robiłem to, co trzeba, nie *przeżywając* tego”? Być może jest to spójny scenariusz, ale nie wydaje się naturalną konstrukcją systemu poznawczego. Gdyby projektowało się taki system, znacznie naturalniejsze byłoby zaprojektowanie go tak, żeby po prostu „widział” różnicę między czerwonym i zielonym, opierał swoje działania bezpośrednio na postrzeżonej różnicy oraz odpowiadał pewnie i bezpośrednio na pytania. W każdym razie ostatnie [rozwiązanie] jest co najmniej *jedną* z możliwych rozsądnych konstrukcji systemu, a to wszystko, czego tu potrzebujemy.

Biorąc pod uwagę tego rodzaju bezpośredni dostęp do stanów informacyjnych, naturalne jest oczekiwanie, że system będzie posługiwał się językiem „przeżyć” i „jakości” do opisu własnego poznawczego punktu widzenia na postrzeganie. Nie jest też zaskakujące, że to wszystko może wydawać się systemowi dość dziwne: owe bezpośrednio znane, niewystawialne stany wydają się tak zasadnicze z punktu widzenia jego dostępu do świata, ale jednocześnie są tak trudne do uchwycenia. W istocie natural-

ne jest przypuszczenie, że systemowi wydawałoby się to dziwne tak samo, jak nam wydaje się dziwna świadomość.

Oto więc zaczątek potencjalnego wyjaśnienia redukcyjnego naszych sądów o świadomości: sądy pojawiają się dlatego, że nasz system przetwarzania zostaje zlokalizowany w przestrzeni informacyjnej z bezpośrednim dostępem do tych lokalizacji, ale z niczym więcej. Ta bezpośrednia wiedza będzie uderzać system jako surowa „jakość”: wie, że stany są różne, ale nie może tego wyrazić inaczej, niż mówiąc, w rezultacie, „jedna z *tych*”. Ten bezpośredni dostęp do surowych różnic prowadzi do sądów o tajemniczym, pierwotnym charakterze tych jakości, o niemożności objaśnienia ich w bardziej podstawowych kategoriach, a także do wielu innych sądów, które często wygłaszamy na temat przeżyć świadomych.

Informacja odgrywa w tym wszystkim zasadniczą rolę. Jest tak dlatego, że system ma dostęp tylko do stanów informacyjnych, że tworzone są różne sądy o surowych „jakościach”. System po prostu wchodzi w różne stany, a późniejsze procesy mają dostęp tylko do struktury różnic tych wcześniejszych stanów, a poza tym do niczego konkretnego. Podstawą działania jest tu system różnic, które czynią różnicę. To informacja i nasz dostęp do niej redukcyjnie wyjaśniają sądy tworzone o świadomości.

Niektórzy mogą tutaj spocząć na laurach, oświadczając, że tajemnica świadomości została usunięta i podano wyjaśnienie. Oczywiście, nie sądzę, by była to prawda: wyjaśnilibyśmy tylko pewne sądy, co jest o wiele prostszą sprawą. Możemy jednak z pożytkiem zastosować zasadę spójności eksplanacyjnej w teorii świadomości. Jeżeli to przede wszystkim stany informacyjne, które są realizowane w trakcie przetwarzania [informacji], odpowiadają za nasze sądy o świadomości, być może te stany informacyjne odpowiadają za samą świadomość.

W rzeczywistości tak właśnie doszedłem do informacyjnego poglądu na świadomość. Jeśli podstawą eksplanacyjną naszych sądów fenomenalnych jest struktura różnic, które czynią różnicę, naturalne jest przypuszczenie, że podstawa eksplanacyjna świadomości może znajdować się w tym samym

miejscu. To wyjaśniałoby, dlaczego nasze sądy są tak dobrze dopasowane do rzeczywistych stanów świadomości. Przeżycie świadome jest realizacją stanu informacyjnego; sąd fenomenalny jest wyjaśniany przez inną realizację tego samego stanu informacyjnego. W pewnym zaś sensie musimy jedynie postulować fenomenalny aspekt informacji, aby mieć pewność, że sądy te będą naprawdę poprawne: naprawdę *istnieje* jakościowy aspekt tych informacji, ujawniający się bezpośrednio w fenomenologii, a nie tylko w systemie sądów. Pozwala to więc świadomości na uzyskanie eleganckiej spójności ze strukturą poznawczą, co prowadzi do ściślejszego powiązania elementów wizji umysłu.

Możemy również zauważyć, że poznawcza rola stanów informacyjnych i epistemologia przeżycia są także dobrze dopasowane. Przeżyciom, które znamy bezpośrednio, odpowiadają fizycznie zrealizowane stany informacyjne, do których system ma bezpośredni (poznawczy) dostęp. System tworzy swoje sądy fenomenalne na podstawie bezpośredniego dostępu do stanów informacyjnych; ten związek przyczynowy odzwierciedla się elegancko w twierdzeniu, że przeżycie – fenomenalna realizacja tego samego stanu informacyjnego – *uzasadnia* tworzone przekonania fenomenalne. Po obu stronach podstawową rolę odgrywa ten sam stan informacyjny; w jednym przypadku jest to po prostu realizacja fizyczna, a w drugim – fenomenalna.

Nie jest to bynajmniej powalający dowód, że informacyjne podejście do świadomości musi być prawdziwe. Zapewnia jednak konkretne podejście, które ma dodatkowe zalety.

#### 4

### czy przeżycia są wszechobecne?

W tym momencie czytelnicy prawdopodobnie zechcą wysunąć zarzut, że informacje są *wszechobecne*. Informacje są wszędzie, nie tylko w systemach, które standardowo uznajemy za świadome. Mój odtwarzacz płyt

kompaktowych realizuje informację; silnik mojego samochodu realizuje informację; nawet termostat realizuje informację. W rzeczywistości przy moim rozumieniu tego pojęcia informacja znajdują się wszędzie tam, gdzie istnieje przyczynowość. /Przyczynowość jest wszędzie, więc wszędzie są też informacje. Z pewnością jednak przeżycia nie są wszędzie?

Zwolennik podejścia opartego na informacji może zareagować na tę sytuację dwojako. Pierwszą i najbardziej oczywistą metodą jest poszukiwanie dodatkowych ograniczeń *rodzaju* informacji, które są istotne dla przeżycia. Z przeżyciem wiąże się nie tyle dowolna fizycznie zrealizowana przestrzeń informacyjna, ile przestrzeń z określonymi własnościami. Metoda ta wymagałaby uważnego rozważenia, jakie mogłyby być dalsze ograniczenia oraz w jaki sposób mogłyby one pasować do praw fundamentalnych. Rozważę utrzymane w tym duchu strategię później, teraz jednak chciałbym zbadać podejście alternatywne. Polega ono na zagryzieniu zębów i przyjęciu, że wszystkie informacje wiążą się z przeżyciem. Jeśli tak, to nie tylko informacje są wszechobecne. Przeżycia też.

Jeśli to prawda, to przeżycia wiążą się też z bardzo prostymi systemami. Idea ta jest często traktowana jako skandaliczna, a nawet szalona. Sądzę jednak, że zasługuje na dokładne zbadanie. Nie jest dla mnie *oczywiste*, że ten pomysł jest poroniony, a w pewnym sensie ma swój urok. Zbadam tutaj zatem powody ewentualnego odrzucenia tego poglądu, aby sprawdzić, czy są one przekonujące, biorąc jednocześnie pod uwagę pozytywne względy skłaniające do potraktowania go poważnie.

### jak to jest być termostatem?

Aby skupić się na przykładzie, weźmy pod uwagę system przetwarzania informacji, który jest niemal maksymalnie prosty: termostat. Termostat, traktowany jako urządzenie do przetwarzania informacji, ma tylko trzy stany informacyjne (jeden prowadzi do chłodzenia, drugi do ogrzewania, a trzeci do braku działań). Tak więc twierdzenie głosi, że każdemu temu stanowi

informacyjnemu odpowiada stan fenomenalny. Te wszystkie trzy stany fenomenalne będą różne, a zmiana stanu informacyjnego zmieni stan fenomenalny. Możemy zapytać: jaki jest charakter tych stanów fenomenalnych? Czyli – jak to jest być termostatem?

Z pewnością bycie termostatem nie będzie bardzo interesujące. Przetwarzanie informacji jest tak proste, że powinniśmy się spodziewać, iż odpowiadające stany fenomenalne będą równie proste. Będziemy mieli do czynienia z trzema pierwotnie różnymi stanami fenomenalnymi, bez dalszej struktury. Zapewne możemy ujmować te stany przez analogię do naszych przeżyć czerni, bieli i szarości: termostat może mieć całe czarne pole fenomenalne, całe białe lub całe szare. W ten sposób jednak imputujemy zbyt wyrafinowaną strukturę przeżyciom termostatu, sugerując wymiarowość pola widzenia oraz stosunkowo bogaty charakter czerni, bieli i szarości. Powinniśmy się spodziewać czegoś naprawdę dużo prostszego, niemającego odpowiednika w naszych przeżyciach. Prawdopodobnie nie będziemy w stanie wyobrazić sobie tych przeżyć lepiej, niż osoba niewidoma jest w stanie wyobrazić sobie wzrok, a człowiek – jak to jest być nietoperzem, ale możemy przynajmniej intelektualnie dowiedzieć się czegoś o ich podstawowej strukturze.

Aby ten pogląd wydawał się mniej ekscentryczny, możemy zastanowić się, co dzieje się z przeżyciami, kiedy schodzimy w dół na skali złożoności. Zaczynamy od znanych przypadków ludzi, u których bardzo złożone przetwarzanie informacji prowadzi do znanych nam złożonych przeżyć. Gdy przechodzimy do systemów mniej skomplikowanych, nie wydaje się, żeby było wiele powodów, by wątpić, że psy – a nawet myszy – są świadome. Niektórzy to kwestionowali, ale myślę, że dzieje się tak często ze względu na zrównywanie świadomości fenomenalnej i samoświadomości. Myszy mogą nie mieć zbyt wielkiego poczucia „ja” i mogą nie mieć introspekcji, ale wydaje się całkiem prawdopodobne, że jest coś takiego jak „jak to jest być myszą”. Myszy postrzegają swoje środowisko przez wzorce przepływu informacji podobne do wzorców w naszych mózgach, choć znacznie mniej skompliko-

wane. Czymś naturalnym jest zatem hipoteza, że „rozmaitości postrzeżeniowej”, którą, jak wiemy, myszy mają, odpowiada „rozmaitość fenomenalna”. Rozmaitość postrzeżeniowa myszy jest dość bogata – mysz może poczynić wiele postrzeżeniowych rozróżnień – więc jej rozmaitość fenomenalna może być również dość bogata. Jest na przykład prawdopodobne, że każdemu rozróżnieniu, którego może dokonać układ wzrokowy myszy i które może zastosować w postrzeganiu środowiska, odpowiada rozróżnienie fenomenalne. Nie można *udowodnić*, że tak jest, ale wydaje się to najnaturalniejszym sposobem pojmowania fenomenologii myszy.

Gdy schodzimy w dół skali – przez jaszczurki i ryby aż do ślimaków – pod uwagę należy brać podobne rzeczy. Nie wydaje się, by należało zakładać, że fenomenologia zanika, gdy utrzymuje się stosunkowo złożona psychologia postrzeżenia. Jeśli tak się dzieje, to albo następuje radykalne zerwanie ciągłości od złożonych przeżyć do żadnych, albo gdzieś po drodze fenomenologia przestaje być zsynchronizowana z postrzeganiem, tak że przez jakiś czas istnieje dość bogata rozmaitość postrzeżeniowa, której towarzyszy znacznie uboższa rozmaitość fenomenalna. Pierwsza hipoteza wydaje się mało prawdopodobna, a druga sugeruje, że pośrednie systemy miałyby życie wewnętrzne dziwnie oderwane od ich zdolności poznawczych. Inna hipoteza jest na pewno przynajmniej tak samo prawdopodobna. Prawdopodobnie o wiele mniej ciekawe jest być rybą niż człowiekiem, z prostszą fenomenologią odpowiadającą jej prostszej psychologii, ale wystarczająco rozsądne wydaje się, że coś tam jest.

Gdy schodzimy na skali od ryb i ślimaków przez proste sieci neuronowe aż do termostatów, w którym momencie świadomość powinna zaniknąć? Fenomenologia ryb i ślimaków prawdopodobnie nie będzie prymitywna, ale stosunkowo złożona, odzwierciedlając różne rozróżnienia, które mogą one poczynić. Zanim fenomenologia całkowicie zaniknie, przypuszczalnie stanie się jakąś maksymalnie prostą fenomenologią. Wydaje mi się, że najnaturalniejszym miejscem, gdzie może to nastąpić, jest system z odpowiednio prostą „psychologią postrzeżenia”, taki jak termostat. Wydaje się,

że termostat realizuje rodzaj przetwarzania informacji obecny u ślimaka lub ryby, ale sprowadzony do jego najprostszej formy, więc być może mógłby również mieć odpowiedni rodzaj fenomenologii w jej najprostszej postaci. Czyni jedno lub dwa istotne rozróżnienia, od których zależy jego działanie; mnie przynajmniej nie wydaje się nierozsądne, że mogłyby się z tym wiązać rozróżnienia przeżyć.

Oczywiście, sprawy *mogłyby* potoczyć się inaczej, gdy schodzimy w dół na skali złożoności – i nie jest to żaden dowód, że termostaty *muszą* mieć przeżycia. Wydaje się jednak rozsądne, że sprawy mogłyby się tak właśnie potoczyć, a po zastanowieniu jest to równie naturalne, jak inne możliwości. Można dyskutować, czy wchodzące tu w grę rozumowanie jest tylko przedłużeniem rozumowania, na którego mocy przypisujemy przeżycia psom lub myszom. Skoro tylko zaczniemy się zastanawiać nad tym, co może się dziać w przeżyciu myszy, i uzasadnimy to jej psychologią postrzeżenia, ekstrapolacja na prostsze systemy zaczyna się przynajmniej wydawać naturalniejsza niż na początku.

Ktoś, kto za „szalone” uznaje założenie, że termostat może mieć przeżycia, jest nam co najmniej winny objaśnienie, *dłaczego* jest to szalone. Przypuszczalnie dlatego, że istnieje własność, której nie ma termostat, a która jest w oczywisty sposób wymagana dla przeżyć; ale żadna taka własność nie wydaje mi się oczywista. Być może istnieje zasadniczy składnik przetwarzania [informacji], którego nie ma termostat, a ma go mysz, albo nie ma go mysz, a posiada ją człowiek, ale nie dostrzegam takiego składnika, który *w oczywisty sposób* jest wymagany dla przeżyć – w istocie nie jest oczywiste, że taki składnik musi istnieć.

Mówiąc, że termostaty mają przeżycia, nie mówimy oczywiście, że mają bogate życie umysłowe. Termostat nie będzie *samoświadomy*; nie będzie w najmniejszym stopniu inteligentny i nie twierdziłbym, że termostat może *myśleć*<sup>3</sup>. Część oporu wobec idei świadomego termostatu może brać się

<sup>3</sup> W sprawie twierdzenia, że termostaty mają przekonania i pragnienia, patrz jednak McCarthy 1979.

ze współwystępowania przeżyć z innymi funkcjami psychicznymi, z których prawie na pewno wszystkie wymagają znacznie większej złożoności. Każda z tych cech ma duży składnik psychologiczny i jest prawdopodobne, że do urzeczywistnienia odpowiednich ról przyczynowych potrzebny byłby skomplikowany system. Po rozróżnieniu jednak własności fenomenalnych i psychologicznych idea świadomego termostatu wydaje się mniej groźna. Musimy jedynie wyobrazić sobie coś w rodzaju nieartykułowanego „błysku” przeżyć bez jakichkolwiek pojęć, myśli ani skomplikowanego przetwarzania [informacji].

Innym powodem odrzucania idei świadomego termostatu przez niektórych jest to, że nie można znaleźć żadnego *miejsca* na świadomość w systemie. On sam wydaje się zbyt prosty, a świadomość wydaje się tu bez żadnej roli do odegrania. Ta reakcja oznacza jednak niezrozumienie poglądu nieredukcyjnego: świadomości nigdy nie znajdzie się w systemie przy bliższym badaniu – zawsze będziemy w stanie zrozumieć przetwarzanie [informacji] bez odwoływania się do świadomości. Jeśli świadomość nie jest logicznie superwenientna, nie powinniśmy oczekiwać, że znajdziemy „miejsce” na świadomość w organizacji systemu; świadomość jest dość odmienna od własności przetwarzania w systemie.

– Możliwe, że niektórzy nie chcą przyjąć możliwości świadomych termostatów po prostu dlatego, że zbyt dobrze *rozumiemy* termostaty. Wiemy wszystko o ich przetwarzaniu i wydaje się, że nie ma żadnego powodu, by przywoływać świadomość. Ale termostaty naprawdę nie różnią się tutaj od mózgow. Nawet kiedy doskonale zrozumiemy przetwarzanie w mózgu, wciąż jeszcze będzie się wydawać, że nie ma powodu, by przywoływać świadomość. Jediną różnicą jest to, że to, co dzieje się w mózgu, jest dziś do tego stopnia tajemnicą, że można ulec pokusie przypuszczenia, iż świadomość jakoś „mieści się” w tych procesach mózgowych, których jeszcze nie rozumiemy. Jak jednak argumentowałem, nawet zrozumienie tych procesów samo nie spowoduje uwzględnienia świadomości; tak więc po raz kolejny mózgi i termostaty są na tym samym poziomie.

Fakt, że da się *zbudować* termostat samodzielnie, nie tworząc w nim żadnej świadomości, może niepokoić. To samo jednak dotyczy oczywiście mózgu, przynajmniej co do zasady. Kiedy budujemy mózg (powiedzmy przez reprodukcję i rozwój), świadomość pojawia się „za darmo”; to samo będzie dotyczyć termostatu. Nie powinniśmy oczekiwać lokalizacji świadomości jako fizycznego składnika systemu! Niektórzy mogą się martwić faktem, że termostat nie jest *żywy*, ale trudno zrozumieć, dlaczego miałoby to być zasadniczą różnicą. Nieucieleśniony mózg krzemowy w rodzaju omówionego w poprzednim rozdziale prawdopodobnie nie zostanie uznany za *żywy*, ale *widzieliśmy*, że może być świadomy. A jeśli argumenty w ostatnim rozdziale są słuszne, to fakt, że termostat nie składa się z *biologicznych* składników, nie robi tu – co do zasady – żadnej różnicy.

Intuicyjny opór może częściowo brać się stąd, że nie wydaje się, aby w termostacie było miejsce na kogoś lub coś *mającego* przeżycia: gdzie w termostacie może zmieścić się *podmiot*? W układach fizycznych nie należy jednak szukać homunkulusa służącego jako podmiot. Podmiotem jest cały system, albo lepiej, podmiot wiąże się z systemem w taki sposób, w jaki wiąże się z mózgiem. Trudno mówić o tym w sposób właściwy. Nie powiedzielibyśmy, że mój *mózg* ma przeżycia, ściśle mówiąc, ale że *ja* je mam. Jakkolwiek rozumiemy tę relację, to samo odnosi się do termostatów: ściśle rzecz biorąc, najlepiej chyba nie mówić, że termostat ma przeżycia (choć nadal będę mówił w taki swobodny sposób), ale że przeżycia są związane z termostatem. Nie znajdziemy podmiotu „wewnątrz” termostatu, tak samo, jak nie znajdziemy go wewnątrz mózgu.

Powróćmy do pozytywnych argumentów na rzecz tego, że proste systemy mają przeżycia: w ten sposób można uniknąć konieczności tego, by świadomość „pojawiła się” na pewnym poziomie złożoności. Jest coś dziwnego w idei, że system z  $n$  elementami mógłby nie być świadomy, ale system z  $n+1$  elementami mógłby. Nie możemy uniknąć podjęcia decyzji – tak samo, jak moglibyśmy uniknąć podjęcia decyzji, kiedy dokładnie ktoś staje się „łysy”: w tym drugim przypadku istnieje wiarygodny stopień se-

mantycznej nieokreśloności, ale jest znacznie mniej wiarygodne, że to, czy system jest świadomy, może być nieokreślone. (Jest tak zwłaszcza wtedy, gdy przyjmujemy pogląd nieredukcyjny, zgodnie z którym nie możemy wyeksplikować faktów dotyczących przeżyć w kategoriach bardziej podstawowych faktów – tak jak eksplikujemy nieokreślone kwestie dotyczące łysienia w kategoriach określonych faktów o liczbie włosów na głowie.) Chociaż *mogłoby* być tak, że przeżycie pojawia się w konkretnym momencie, każdy konkretny moment wydaje się arbitralny, więc teoria, która pozwala uniknąć konieczności podejmowania tej decyzji, zyskuje pewną prostotę.

Oto ostatnia okoliczność będąca świadectwem tego, że proste systemy mają przeżycia: jeśli przeżycie jest naprawdę własnością fundamentalną, wydaje się naturalne, że jest ona powszechna. Na pewno wszystkie inne właściwości fundamentalne, o których wiemy, występują nawet w prostych systemach i w całym wszechświecie. Byłoby dziwne, gdyby własność fundamentalna była egzemplifikowana po raz pierwszy stosunkowo późno w historii wszechświata i to tylko w rzadkich złożonych systemach. Nie ma *sprzeczności* w idei, że własność fundamentalna ma być egzemplifikowana tylko od czasu do czasu; w normalnych warunkach jednak alternatywne ujęcie wydaje się bardziej wiarygodne.

### quo vadis, panpsychizmie?

Jeśli istnieją przeżycia związane z termostatem, to przeżycia są zapewne *wszędzie*: wszędzie tam, gdzie istnieje oddziaływanie przyczynowe, tam jest informacja, a gdzie jest informacja, tam jest przeżycie. (Stany informacyjne można znaleźć w kamieniu – gdy na przykład rozszerza się i kurczy – czy nawet w różnych stanach elektronu. Jeśli więc nieograniczona zasada podwójnego aspektu jest poprawna, będą istnieć przeżycia związane z kamieniem lub elektronem.)

(Nie powiedziałbym właściwie, że kamień *ma* przeżycia lub że kamień *jest świadomy* w sposób, w jaki mógłbym luźno powiedzieć, że termostat

ma przeżycia lub jest świadomy. Kamień, w przeciwieństwie do termostatu, nie jest wyróżniony jako system przetwarzania informacji. Jest to po prostu wyróżniony jako przedmiot, więc związek z przeżyciem jest mniej bezpośredni. Może lepiej byłoby powiedzieć, że kamień zawiera systemy, które są świadome: prawdopodobnie istnieje wiele takich podsystemów, których przeżycia nigdy nie byłyby uznawane kanonicznie za przeżycia kamienia (nie bardziej niż moje przeżycia są uznawane za przeżycia mojego biura). W wypadku termostatu natomiast istnieje kanoniczna, związana z nim przestrzeń informacyjna, dlatego bardziej rozsądne wydaje się mówienie o kanonicznych przeżyciach termostatu. Oczywiście nawet taki sposób użycia tego określenia jest dość luźny, jak wspomniano powyżej.)

Pogląd, że przeżycia są wszędzie tam, gdzie istnieje oddziaływanie przyczynowe, pozostaje sprzeczny z intuicją. Jest to jednak pogląd, który po zastanowieniu może stać się zaskakująco satysfakcjonujący, gdyż lepiej integruje świadomość z porządkiem przyrody. Jeśli jest słuszny, świadomość nie pojawia się nagle i skokowo, z izolowanymi złożonymi systemami arbitralnie wytwarzającymi bogate przeżycia świadome. Jest to raczej bardziej jednolita własność wszechświata, gdzie bardzo proste systemy mają bardzo prostą fenomenologię, a systemy złożone – złożoną fenomenologię. To sprawia, że świadomość jest w pewnym sensie mniej „szczególna”, a więc bardziej rozsądna.

Ciekawe wydaje się pytanie, czy *aktywna* przyczynowość jest wymagana dla przeżyć. Czy termostat może mieć przeżycia, gdy pozostaje w stałym stanie (w pewnym sensie „powoduje” wyjście, ale nie *robi* naprawdę nic)? Czy też ma przeżycia tylko wtedy, gdy się zmienia? Przyczynowość leżąca u podstaw przeżyć w mózgu wydaje się w większości aktywna, gdyż odpowiednie informacje są stale przetwarzane, neurony są aktywowane i tak dalej. Z drugiej strony, możliwe, że rozróżnienia aktywnej i pasywnej przyczynowości nie można dokonać na poziomie fundamentalnym, w którym to przypadku oba [rodzaje] mogłyby być traktowane jednakowo. Nie znam

odpowiedzi na to pytanie, ale mam intuicję, że pewnego rodzaju aktywność jest wymagana dla przeżyć.

Jedną z możliwości, której do tej pory nie rozważałem, ale której nie można wykluczyć, jest to, że proste systemy nie mają własności fenomenalnych, ale mają własności protofenomenalne. W rozdziale IV wspomniałem o możliwości, że mogą istnieć własności bardziej fundamentalne od własności fenomenalnych, z których te ostatnie będą się składać. Jeśli rzeczywiście takie własności istnieją, to wydawałoby się naturalne, że są one egzemplifikowane w prostych systemach. Jeśli tak, to termostaty mogłyby nie mieć przeżyć w zwykłym sensie, ale zamiast tego egzemplifikować pokrewnego rodzaju własność, której nie rozumiemy w pełni (może coś w rodzaju protoprzeżycia). W ten sposób zachowalibyśmy jednolity obraz porządku przyrody, wspomniany wyżej, a także moglibyśmy poradzić sobie z problemem „zaniku” (jeśli własności protofenomenalne są fundamentalne, to przeżycia ukonstytuowane z tych własności mogą jednak się stopniowo „pojawiać”). Dzięki temu, że nie twierdzi się tu, iż termostaty mają pełnoprawne przeżycia, pogląd ten może wydawać się nieco mniej „szalony” niż pogląd alternatywny. Oczywiście, ceną tego jest postulowanie klasy nieznanych własności, których nie rozumiemy; ale ta możliwość musi pozostać otwarta.

Tak czy inaczej, pogląd ten ma wiele wspólnego z tym, co jest często określane jako panpsychizm – koncepcja, że wszystko ma umysł. Istnieje kilka powodów, dla których na ogół nie używam tego pojęcia: (1) ponieważ uważam, że posiadanie przeżyć może nie wystarczać do posiadania umysłu w zwykłym sensie, choć może kwalifikować się jako umysł w jego najprostszej formie; (2) ponieważ własności protofenomenalne mogą być nawet bardziej oddalone od zwykłego pojęcia „umysłu”; (3) ponieważ nie uważam, że ściśle poprawne – z powodów wymienionych wyżej – jest powiedzenie, iż kamienie (na przykład) mają przeżycia, chociaż kamienie mogą mieć związane ze sobą przeżycia. Być może podstawowym powodem, dlaczego termin [„panpsychizm”] pozostaje mylący, jest jednak to, że

sugeruje on pogląd, zgodnie z którym przeżycia w prostych systemach takich jak atomy są fundamentalne, a złożone przeżycia są w pewien sposób sumą tych prostszych przeżyć. Choćby sprawy mogłyby również wyglądać w ten sposób, nie ma powodu, by tak musiało być: złożone przeżycia mogą być bardziej samodzielne, niż się tu sugeruje. W szczególności pogląd informacyjny sugeruje obraz, w którym złożone przeżycia są wyznaczane bardziej holistycznie niż tu.

Uwzględniając te zastrzeżenia, trafne jest chyba stwierdzenie, że pogląd stanowi odmianę panpsychizmu. Chciałbym jednak zauważyć, że panpsychizm nie jest metafizycznym fundamentem mojego poglądu: ten fundament to raczej naturalistyczny dualizm z prawami psychofizycznymi. Panpsychizm jest po prostu jednym z możliwych sposobów zachodzenia superwencji przyrodniczej na tym, co fizyczne. W pewnym sensie superwencja przyrodnicza zapewnia *ramy*, a panpsychizm to tylko jeden ze sposobów wypracowania *szczegółów*.

Osobiście jestem znacznie bardziej pewny naturalistycznego dualizmu niż panpsychizmu. Ta ostatnia kwestia wydaje się w ogromnej mierze otwarta. Mam jednak nadzieję, że powiedziałem wystarczająco dużo dla wykazania, że poważnie powinniśmy wziąć możliwość jakiegoś rodzaju panpsychizmu: wydaje się, że nie ma żadnych powalających argumentów przeciwko temu pogładowi, a są różne pozytywne powody, dla których można by go przyjąć.

### **ograniczanie zasady podwójnego aspektu**

Nawet jeśli ktoś jest gotów przyjąć że bardzo proste systemy mają przeżycia, to z ideą, że wszystkie informacje wiążą się z przeżyciami, może mu być nadal nieswojo. Na przykład, wydaje się, że tylko niewielkiej części informacji w przetwarzaniu poznawczym u człowieka odpowiadają informacje w przeżyciach świadomych. Czyż nie jest po prostu faktem, że przetwarzanie informacji jest u nas w większości nieświadome?

Jeśli nieograniczona zasada podwójnego aspektu jest poprawna, to zapewne odpowiedzią na to będzie stwierdzenie, że wszystkie „nieświadome” informacje są realizowane w przeżyciach – tyle że nie są realizowane w *moich* przeżyciach. Na przykład, jeśli istnieje przeżycie związane z jednym z moich neuronów w sposób, w jaki istnieje przeżycie związane z termostatem, to nie oczekiwilibyśmy, że będzie ono należeć do *moich* przeżyć – tak samo, jak nie oczekiwilibyśmy, że moje przeżycie zostanie radykalnie przekształcone, jeśli neuron zostanie zastąpiony małym świadomym homunkulusem. Podobnie może istnieć przeżycie związane z różnymi „nieświadomymi” podsystemami przetwarzania informacji w mózgu – tyle tylko, że te przeżycia należą do innego podmiotu. Istnieje wiele różnych systemów przetwarzania informacji w mózgu, a ten, który jest *moim* odpowiednikiem – być może system udostępniający pewne informacje do pewnego rodzaju globalnej kontroli i relacjonowania – jest tylko jednym z nich. Nie oczekiwałbym, że będę miał bezpośredni dostęp do przeżyć innych systemów, tak samo, jak nie oczekiwałbym, że będę miał bezpośredni dostęp do przeżyć innych ludzi.

Można by się także martwić wszystkimi stosunkowo złożonymi systemami przetwarzania informacji na całym świecie, od mojego odtwarzacza płyt kompaktowych do mojego brzucha. Czy wszystkie one liczą się jako świadome jednostki ze złożonymi przeżyciami? W odpowiedzi warto zauważyć, że systemy te nie mają niczego podobnego do spójnej struktury poznawczej naszego własnego systemu, więc wszelkie ewentualne związane z nimi przeżycia raczej nie będą podobne do naszych. Jeśli na przykład z odtwarzaczem płyt kompaktowych wiążą się przeżycia, to prawdopodobnie nie są niczym więcej niż „płaską” strukturą bitów; a jeśli informacje w moim brzuchu są związane z przeżyciami, to nie ma powodu, by sądzić, że temu przeżyciu odpowiadałoby coś, co uznajemy za umysł. Rodzaje przeżyć, które mamy, powstają tylko wtedy, gdy systemy przetwarzania informacji zostały ukształtowane przez ewolucję w taki sposób, aby miały złożone, spójne struktury poznawcze odzwierciedlające bogatą reprezentację świata zewnętrznego. Jest prawdopodobne, że jedynie bardzo ograniczona grupa

podmiotów przeżyć miałyby strukturę psychologiczną wymaganą do tego, aby potraktować je jako *podmioty działające* (*agents*) lub *osoby*.

To wielkie rozpowszechnienie przeżyć, a zwłaszcza ich rozpowszechnienie w jednym mózgu, może być jednak przyczyną dyskomfortu. Problem potęguje obserwacja, że w przypadku danej przestrzeni informacyjnej łatwo jest zwykle znaleźć wiele nieco innych przestrzeni informacyjnych, po prostu inaczej indywidualizując odpowiednią ścieżkę przyczynową lub dzieląc istotne efekty („różnice”, jakie czyni informacja) w nieco inny sposób. Czy mamy sądzić, że istnieją różne zbiory przeżyć unoszące się nad wszystkimi tymi przestrzeniami informacyjnymi? Jeśli tak, to moglibyśmy mieć wiele bardzo bliskich, ale nieco innych fenomenalnych krewnych powstających z procesów w mojej głowie!

Innym rozwiązaniem jest takie *ograniczenie* zasady podwójnego aspektu, aby zawężyła klasę fizycznie zrealizowanych przestrzeni informacyjnych, które mają fenomenalne odpowiedniki. Najnaturalniejszą strategią może być ograniczenie *sposobu* przetwarzania informacji. W końcu już powiedziałem, że informacje w moim systemie, które odpowiadają najbardziej bezpośrednio moim przeżyciom, to informacje, które są bezpośrednio dostępne dla kontroli globalnej. Jest bardzo mało prawdopodobne, że to „kryterium” w obecnej postaci odegra rolę w prawie fundamentalnym, ponieważ jest pojęciem zbyt nieostrym i należącym do zbyt wysokiego poziomu; w istocie możemy skorzystać z tej zasady tylko wtedy, gdy już zindywidualizowaliśmy system wysokiego poziomu taki jak osoba lub mózg. Może jednak istnieje precyzyjniejsze, prostsze kryterium, które mogłoby się tu sprawdzić.

Możliwe, że podstawowe znaczenie ma *wzmocnienie* informacji. Fizycznie zrealizowane informacje są również realizowane w przeżyciu tylko wtedy, gdy informacja jest *wzmocniona* w określony sposób, stając się dostępną i czyniąc znaczną różnicę na niektórych ścieżkach przyczynowych. Być może można by nawet powiedzieć, że intensywności przeżycia odpowiada stopień wzmocnienia lub coś podobnego. Mogłoby to ładnie pasować do kryterium dostępności globalnej, chociaż może rodzić inne problemy: na przykład

istnieje mnóstwo wzmocnionych informacji, które nie są intuicyjnie świadome; nie jest też oczywiste, jak pojęcie wzmocnienia można sprecyzować.

Inną możliwość stanowi ograniczenie rodzaju *przyczynowości* występującej w systemie. Widzieliśmy, że tam, gdzie jest przyczynowość, tam jest informacja; ale być może tylko pewnego rodzaju przyczynowość liczy się przy indywidualizacji przestrzeni informacyjnych leżących u podstaw przeżyć. Może na przykład istotne są tylko pewne rodzaje „aktywnych” związków przyczynowych, a może wymagane są pewne rodzaje „naturalnych” związków przyczynowych. Istnieje intuicja, że wiele przestrzeni informacyjnych, które można znaleźć w zależności od kryteriów podanych do tej pory, jest w pewnym sensie nienaturalnych; być może istnieje sposób eksplikacji odpowiedniego ograniczenia. Prawdopodobnie i tak dopuszczaloby to bardzo szeroką klasę stanów informacyjnych, ale mogłoby zapobiec astronomicznemu rozrostowi ich liczby.

Nie jestem pewien, jaki powinno wyglądać takie odpowiednie kryterium ograniczające, ale nie znaczy to, że nie może ono istnieć. Możliwe nawet, że mogłoby ono zawęzić odpowiednie przestrzenie informacyjne, tak by nie kwalifikowały się tu informacje w prostych systemach takich jak termostaty. Moja własna intuicja mówi, że nawet mimo ograniczenia zasady podwójnego aspektu informacje w prostych systemach takich jak termostaty mogłyby się i tak kwalifikować. Moim zdaniem, rozpowszechnianie się wielu pokrewnych przeżyć w mózgu wydaje się bardziej sprzeczne z intuicją niż obecność przeżyć w prostych systemach, choć sprawa nie jest przesądzona. W każdym razie istnieje wiele różnych możliwości rozwoju prototeorii w stronę teorii.

## 5 metafizyka informacji

Pozostaje problem tego, jak rozumieć ontologię poglądu [głoszącego] podwójny aspekt informacji i jak dalece poważnie traktować pojęcia stanów

i przestrzeni informacyjnych: czy są to tylko przydatne konstrukty, czy też w jakiś sposób są one ontologicznie fundamentalne? Czy informacja jest pierwotna, czy też rzeczywiście pierwotne jest to, co fenomenalne, i to, co fizyczne, a informacja stanowi jedynie przydatny łącznik?

Można to wszystko rozumieć różnorako. Najprostsze i najmniej ryzykowne jest uznanie fizycznej i fenomenalnej realizacji informacji za całkowicie odrębne cechy, bez żadnego ontologicznego łącznika ponad i poza powiązaniem opartym na prawie i strukturalnym izomorfizmem pewnego rodzaju. Zgodnie z tym poglądem ontologia pozostaje prostą ontologią dualizmu własności, z własnościami fizycznymi, osobnymi własnościami fenomenalnymi i opartym na prawie związku między oboma. Tutaj pojęcie „podwójnego aspektu” należy brać w sposób deflacyjny: jest to tylko barwny sposób opisu dwóch różnych rodzajów skorelowanych własności o podobnej strukturze. A informacja jest po prostu użytecznym narzędziem opisu tej wspólnej struktury, nie odpowiada natomiast niczemu ontologicznie „głębokiemu”.

Możliwe, że to całkowicie adekwatne spojrzenie na rzeczywistość, ale są też bardziej interesujące możliwości. Większość z nich sprowadza się do poważniejszego potraktowania roli informacji. Rozważę jedną z tych możliwości poniżej. Ostrzegam czytelnika, że dyskusja wchodzi głęboko w sferę metafizyki spekulatywnej, ale metafizyki tej prawdopodobnie nie da się uniknąć w przypadku ontologii świadomości.

### **bit tworzy byt**

W fizyce czasami sugeruje się, że informacja jest fundamentalna dla fizycznej struktury wszechświata, a nawet że własności i prawa fizyczne mogą być pochodne od własności i praw informacyjnych. Taki pogląd, że „bit tworzy byt” proponują Wheeler (1989, 1990) i Fredkin (1990); analizują go też antologie Zurka (1990) i Matzkego (1992, 1994). Jeśli to prawda, być może będziemy w stanie przyznać informacji poważniejszą rolę w naszej

ontologii. Aby lepiej to uchwycić, rozważę jeden ważny sposób możliwego postrzegania informacji jako fundamentalnej dla fizyki. To nie jedyna możliwa postać idei „bitu tworzącego byt” (różni się zwłaszcza nieco od poglądu Wheelera)<sup>4</sup>, ale uderza mnie jako chyba najnaturalniejszy sposób rozumienia pojęcia informacji. Interpretacja ta, jak się przekonamy, jest ściśle związana z „russelowskimi” ideami omawianymi w rozdziale IV (s. 265–270).

Takie podejście wyłania się z obserwacji, że w teoriach fizycznych fundamentalne stany fizyczne są efektywnie indywidualizowane jako *stany informacyjne*. Kiedy patrzymy na taką cechę, jak masa lub ładunek, po prostu znajdujemy surową przestrzeń różnic, które czynią różnicę. Fizyka nie mówi nam nic o tym, czym *jest* masa ani czym *jest* ładunek: po prostu podaje nam zakres różnych wartości, które te cechy mogą przyjmować, a także mówi nam, jakie są ich efekty w odniesieniu do innych cech. Jeśli chodzi o teorie fizyczne, specyficzne stany masy lub ładunku mogłyby równie dobrze być czystymi stanami informacyjnymi: liczy się tylko ich położenie w przestrzeni informacyjnej.

Znajduje to odzwierciedlenie w fakcie, że fizyka nie zobowiązuje do żadnego określonego sposobu *realizacji* tych stanów. Jakakolwiek realizacja tych stanów informacyjnych równie dobrze posłuży teorii fizycznej, dopóki zachowuje poprawną strukturę relacji przyczynowych lub dynamicznych między stanami. W końcu, dopóki kształt tych relacji jest ten sam, fizyka będzie *wyglądać* tak samo dla naszych systemów postrzeżeniowych: nie mamy dostępu do żadnych dalszych własności realizacji w świecie zewnętrznym, ponad i poza kształtem sieci przyczynowej. (Z jednym być może wyjątkiem, gdy nasze własności fenomenalne są bezpośrednio związane z własnościami realizującymi.)

Czasami sugerowano nawet, że wszechświat mógłby być gigantycznym komputerem. Fredkin (1990) sugerował, że wszechświat mógłby być

4 Podejście Wheelera koncentruje się na wynikach pomiaru lub „odpowiedziach na pytania tak-nie” jako podstawie wszystkiego, a jako takie może być bliższe pewnej postaci idealizmu niż pogląd, który tutaj przedstawiam.

ogromnym automatem komórkowym, realizowanym na najniższym poziomie przez ogromną strukturę bitów<sup>5</sup>. Leckey (1993) zasugerował, że cała czasoprzestrzeń mogłaby być oparta na procesie obliczeniowym, z osobnymi rejestrami dla każdej egzemplifikowanej cechy fundamentalnej świata. Dopóki między tymi rejestrami istnieją odpowiednie związki przyczynowe, żadna z istot na tym świecie nie stałaby się też przez to mądrzejsza. Przykład komputera pokazuje, jak duży jest zakres możliwych sposobów realizacji obiektów fizycznych, które „znamy”, dopóki istnieją obiekty, które odgrywają odpowiednie role przyczynowe. To należy więc do „metafizyki fizyki”: spekulacji na temat ontologii leżącej u podstaw struktury przyczynowej samej czasoprzestrzeni.

Ten rodzaj metafizyki nie jest oczywiście czymś, czym zajmuje się sama fizyka. Może ona pozostać zupełnie obojętna na pytania o sposób realizacji jej cech, a nawet o to, czy cechy są „realizowane” w ogóle w jakikolwiek podobny sposób. Jeśli chodzi o fizykę, stan świata mógłby równie dobrze wyczerpywać się w charakterystyce informacyjnej. Jeśli są jakieś dalsze własności „realizujące” niższego poziomu, nie odgrywają żadnej bezpośredniej roli w teoriach fizycznych. Kuszące byłoby pozbycie się ich w ogóle.

Prowadziłoby to do obrazu świata jako świata czystej informacji. Każdej fundamentalnej jego cesze odpowiada przestrzeń informacyjna, a gdziekolwiek fizyka uznaje te cechy za egzemplifikowane, egzemplifikowany jest stan informacyjny z odpowiedniej przestrzeni. Dopóki między tymi stanami informacyjnymi zachodzą właściwe relacje, wszystko będzie, jak należy. Zgodnie z tym obrazem świata nie ma nic więcej do powiedzenia. Informacja jest wszystkim, co istnieje.

Tak rozumiem koncepcję świata „bit tworzy byt”. To osobliwie piękna koncepcja: obraz świata jako czystego strumienia informacyjnego, bez żadnej dodatkowej substancji. (Niektóre wersje poglądu mogą także dopuszczać *czasoprzestrzeń* jako pierwotne ramy, w których osadzone są

<sup>5</sup> Patrz także ciekawe ujęcie idei Fredkina u Wrighta (1988), aby dowiedzieć się więcej o metafizyce leżącej u podstaw jego poglądów.

przestrzenie informacyjne; inne wersje uznają czasoprzestrzeń za konstytuowaną przez relacje między przestrzeniami informacyjnymi.) Świat jest po prostu światem pierwotnych różnic oraz przyczynowych i dynamicznych relacji między tymi różnicami. Zgodnie z tym poglądem próba powiedzenia czegoś więcej o świecie jest błędem.

### ugruntowanie informacji w fenomenologii

Wydaje się, że z tym obrazem świata są dwa podstawowe problemy. Pierwszy z nich stanowi sama świadomość. Wydaje się, że w jej przypadku mamy coś poza i ponad czystą przestrzeń informacyjną. Własności fenomenalne mają naturę wewnętrzną, której nie wyczerpuje całkowicie ich położenie w przestrzeni informacyjnej; tymczasem wydaje się, że czysto informacyjne widzenie świata nie pozostawia miejsca na te wewnętrzne jakości.

Drugi problem polega na tym, że pojęcie czystego strumienia informacyjnego nie jest w sposób oczywisty spójne. Można przeczuwać, że zgodnie z tym poglądem świat ma za mało substancji, aby *być* światem. Czy mogłyby istnieć różnice, które są różnicami *pierwotnymi*, nieugruntowanymi w różnicach żadnych jakości niższego poziomu? Można uznawać za prawdopodobne, że każda konkretna różnica w świecie musi być ugruntowana, czyli, że *musi istnieć różnica w czymś*.

Ten problem jest ściśle związany z problemami, jakie pojawiają się w przypadku poglądu o „czystym strumieniu przyczynowym” omówionym w rozdziale IV (s. 265–270), którego obecnie prezentowany pogląd jest wariantem. Tamten pogląd usuwał ze świata wszystkie jakości wewnętrzne, pozostawiając świat związków przyczynowych, pozbawiony, jak się zdawało, czegokolwiek, co samo mogłoby być przyczyną. Aktualny pogląd może wypaść nieco lepiej, dopuszczając stany informacyjne jako człony relacji przyczynowych, ale stany te są wyjątkowo niesubstancjalne, będąc jedynie *różne* od siebie i nie mając żadnego własnego charakteru. Można sądzić, że ten obraz świata bez charakteru wewnętrznego w ogóle nie jest obrazem świata.

W istocie można argumentować, że przestrzenie informacyjne *muszą* mieć jakiś dodatkowy dalszy charakter. Możliwe, że dwie fundamentalne własności będą miały ten sam rodzaj struktury informacyjnej – obie na przykład wymagają rzeczywistych wartości z kontinuum. Jeśli fizyka jest czystą informacją, to nic nie pozwoli odróżnić egzemplifikacji obu przestrzeni informacyjnych. Musi jednak istnieć *jakaś* różnica między nimi, gdyż dwie własności pojawiają się w zupełnie innych prawach i mają różny efekt w odniesieniu do innych cech świata. Musi więc istnieć coś dalszego, aby można było rozróżnić te egzemplifikacje; coś, co wykracza poza czystą informację. Wydawałoby się, że do ich rozróżnienia jest potrzebna jakiegoś rodzaju jakość wewnętrzna.

Z tymi problemami można próbować radzić sobie w różny sposób. Można zdecydować, że drugi problem nie jest do końca istotny i zadowolili się fizyką czystej informacji, a następnie można spróbować uwzględnić własności fenomenalne jako związane w jakiś sposób z tą informacją na mocy praw. Na drugi problem można ewentualnie odpowiedzieć, postulując własności wewnętrzne, na których ugruntowane są fizyczne przestrzenie informacyjne, a pierwszy problem rozwiązywać, wprowadzając osobno własności fenomenalne.

Najbardziej intrygującą strategią jest jednak próba jednoczesnej odpowiedzi na oba problemy. Pierwszy problem zawiera sugestię, że mamy bezpośrednią wiedzę na temat pewnej wewnętrznej natury w świecie – poza i ponad czystą informacją – w postaci własności fenomenalnych; a drugi sugeruje z kolei, że możemy *potrzebować* pewnej wewnętrznej natury w świecie, aby ugruntować stany informacyjne. Może zatem wewnętrzna natura wymagana do ugruntowania stanów informacyjnych jest ściśle związana z wewnętrzną naturą występującą w fenomenologii. Być może jedna jest nawet konstytutywna dla drugiej. W ten sposób możemy uzyskać tanią i elegancką ontologię oraz rozwiązać dwa problemy za jednym zamachem.

Po raz kolejny jest to ściśle związane z sugestią russellofską opisaną w rozdziale IV, zgodnie z którą nieznanne własności wewnętrzne same

są uznawane za własności fenomenalne (lub protofenomenalne). . . . .  
potrzebował tych własności jako leżących u podstaw związków przyczynowych podawanych w fizyce, a my potrzebujemy ich do ugruntowania stanów informacyjnych (różnic, które czynią różnicę) postulowanych przez fizykę. To w istocie ten sam problem. W obu przypadkach mamy poczucie, że dostajemy dwa rozwiązania w cenie jednego. Potrzebujemy pewnych własności wewnętrznych, aby zrozumieć świat fizyczny, i musimy znaleźć miejsce na własności wewnętrzne ujawniające się w fenomenologii. Dwa problemy wydają się dobrze dopasowane.

Tak więc sugestia polega na tym, że przestrzenie informacyjne wymagane przez fizykę są same w sobie ugruntowane we własnościach fenomenalnych lub protofenomenalnych. Każda egzemplifikacja takiej przestrzeni informacyjnej jest w rzeczywistości realizacją *fenomenalną* (lub protofenomenalną). Ilekroć realizowana jest cecha taka jak masa i ładunek, stoi za nią własność wewnętrzna: własność fenomenalna lub protofenomenalna, lub, mówiąc skrótowo, własność *mikrofenomenalna*. Będziemy mieli zbiór podstawowych przestrzeni mikrofenomenalnych, po jednej dla każdej fundamentalnej własności fizycznej, a te przestrzenie będą ugruntowywać przestrzenie informacyjne wymagane przez fizykę. Ostateczne różnice są różnicami mikrofenomenalnymi.

Oczywiście pogląd ten znowu wymaga pewnej odmiany „skandalicznego” panpsychizmu, ale argumentowałem już, że ten panpsychizm nie jest aż tak nierozsądny, jak powszechnie przypuszczano. Skoro sugerowałem, że mogą istnieć własności fenomenalne tam, gdzie jest informacja, również dobrze możemy zaprząć te własności do odegrania użytecznej roli.

Ontologię, do jakiej w ten sposób dochodzimy, można naprawdę nazwać ontologią podwójnego aspektu. Fizyka wymaga stanów informacyjnych, ale dba tylko o ich relacje, a nie naturę wewnętrzną; fenomenologia wymaga stanów informacyjnych, ale troszczy się tylko o naturę wewnętrzną. Mój pogląd postuluje istnienie jednego podstawowego zbioru stanów informacyjnych łączącego obie [dziedziny]. Można powiedzieć, że wewnętrzne

aspekty tych stanów są fenomenalne, a zewnętrzne aspekty – fizyczne. Albo w postaci hasła: przeżycie to informacje od wewnątrz; fizyka to informacje z zewnątrz.

### co z fenomenologią makroskopową?

Wszystko to działa bardzo dobrze jako ontologia, choć z pewnością dosyć dzika. Zanim jednak damy się zbyt ponieść, powstaje zasadnicze pytanie: jak można uzgodnić tę ontologię ze szczegółami *teorii* psychofizycznej? W szczególności, jak uzgodnić ją z psychofizycznymi prawidłowościami na poziomie makroskopowym? Problem polega na tym, że zasada podwójnego aspektu znajduje tutaj zastosowanie na fundamentalnym poziomie fizycznym z fizycznie zrealizowanymi informacjami *mikroskopowymi* mającymi realizację fenomenalną. Dla teorii świadomości potrzebujemy *makroskopowej* fizycznie zrealizowanej informacji, aby mieć także realizację fenomenalną. A wcale nie jest oczywiste, że tego typu „fenomenologię makroskopową” można wyprowadzić z fenomenologii mikroskopowej.

Na pierwszy rzut oka nasze przeżycia świadome nie wydają się na przykład żadnego rodzaju sumą własności mikrofenomenalnych odpowiadających fundamentalnym cechom fizycznym w mózgu. Nasze przeżycia wydają się znacznie bardziej holistyczne i znacznie bardziej jednorodne niż prosta suma. Jest to wersja „problemu ziarna” podniesionego przez Sellarsa (1965) jako problem dla materializmu: jak przeżycie może być identyczne z ogromnym zbiorem zdarzeń fizjologicznych, biorąc pod uwagę jednorodność pierwszego i „drobnoziarnistość” tych ostatnich? Analogiczny problem jest szczególnie palący dla tego rodzaju poglądów russellovskich, które tu omawiam<sup>6</sup>. Jeśli rdzeń fenomenologii wyczerpuje się w mikrofenomenologii, to trudno sobie wyobrazić, jak można byłoby wyprowadzić jednolitą, ustrukturyzowaną fenomenologię makroskopową;

<sup>6</sup> Te problemy w kontekście poglądu russellovskiego inspirująco omawiają Foster (1991, s. 119–130) i Lockwood (1992).

możemy oczekiwać raczej „postrzępionej”, nieustrukturyzowanej zbieraliny fenomenalnej.

Można radzić sobie z tym na różne sposoby. Po pierwsze, można spróbować postawić sprawę tak, że ontologia dwuaspektowa obowiązuje na wszystkich poziomach, nie tylko na poziomie mikroskopowym. Oznacza to, że nawet fizyczne przestrzenie informacyjne na poziomie makroskopowym są ugruntowane w realizacji fenomenalnej. Można argumentować, że poziom mikroskopowy nie jest bynajmniej uporządkowany: obiekty są tam prostsze, ale nie musi on być szczególnie pod względem ontologicznym. Argumenty podane na rzecz postrzegania świata fizycznego w kategoriach informacyjnych stosuje się również na poziomie makroskopowym. Można argumentować, że nawet na tym poziomie różnicę czyni właśnie przestrzeń różnic makroskopowych, z których każda mogłaby być zrealizowana w odpowiedniej fenomenologii.

Sęk w tym, że może nie być miejsca na wszystkie te odrębne realizacje fenomenalne. Gdy mamy fundamentalne cechy fizyczne zrealizowane w fenomenalnych przestrzeniach informacyjnych, informacje makroskopowe wydają się już ugruntowane: różnice, które czynią różnicę, są teraz ugruntowane w konfiguracjach mikroskopowych cech fizycznych, które same są ugruntowane w mikrofenomenologii. Mimo to można byłoby próbować wprowadzać odrębne fenomenalne ugruntowanie, ale wydaje się to zbędne i teoretycznie mniej eleganckie niż analogiczne posunięcie w przypadku poziomu mikroskopowego. Można byłoby próbować usunąć tę zbędność, stanowiąc, że ugruntowanie makroskopowe jest *pierwotne*, ale wówczas trudno byłoby radzić sobie z przypadkami izolowanych systemów mikroskopowych i tym podobnych. Nie jest więc jasne, że „gruntujące” podejście do ontologii dwuaspektowej może być zastosowane bezpośrednio na poziomie makroskopowym.

Po drugie, można byłoby spróbować zrozumieć, w jaki sposób makroskopowa fenomenologia może być *konstituowana* przez te własności mikrofenomenalne. Na pierwszy rzut oka nie wydaje się ona żadną prostą

sumą ani zbiorowiskiem tych własności – to prowadziłoby bezpośrednio do problemów z „postrzępieniem”. Jednak być może problem polega tylko na tym, że nie rozumiemy umysłowej relacji część-całość, jak ujął to Nagel (1997). To znaczy, że nie mamy adekwatnej koncepcji „sumowania” niskopoziomowych własności mikrofenomenalnych do wysokopoziomowej fenomenologii. Myślimy o tym w kategoriach analogii fizycznej, opartej na sumowaniu mikrofizyki do makrofizyki, ale może to być błędnym sposobem myślenia. Być może fenomenologia jest konstytuowana zupełnie inaczej.

Na przykład mogłoby się okazać, że własności mikrofenomenalne sumują się do makrofenomenologii w sposób odzwierciedlający ich wspólną strukturę *informacyjną*, a nie ich wspólną strukturę czasoprzestrzenną. Jeśli zbiorowisko tych własności wspólnie realizuje złożony stan informacyjny na mocy związków przyczynowych między nimi, być może moglibyśmy spodziewać się, że wszelka pochodna makrofenomenologia będzie miała postać tego stanu informacyjnego. W końcu najważniejszą rolą własności mikrofenomenalnych jest realizacja stanów informacyjnych, więc nie byłoby całkowicie zaskakujące, że struktura informacyjna odgrywa rolę w relacjach konstytutywnych między własnościami. Gdyby tak było, to wszelkie pochodne stany makrofenomenalne miałyby „gładką” strukturę informacyjną, którą przewiduje pierwotna zasada podwójnego aspektu. Nie jest to łatwe do zrozumienia, ale przecież nie możemy oczekiwać, że codzienne rozumienie dziedziny fizycznej będzie miało zastosowanie do dziedziny fenomenalnej. Może więc być tak, że lepsze zrozumienie natury samej fenomenologii byłoby zgodne z tym poglądem na jego konstytucję.

Gdyby okazało się, że żadna relacja konstytucji nie może tak działać, możemy wypróbować trzeci wariant, którym jest powiązanie makrofenomenologii z mikrofenomenologią za pomocą *praw*<sup>7</sup>. Na przykład, mogłoby

<sup>7</sup> Lockwood (1992) sugeruje, że w ramach poglądu russellovskiego można w tym celu przywołać surowe prawa. Nie odnosi się on do zarzutu, że wprowadzenie dalszych praw w tym charakterze wydaje się narażać na szwank pierwotne zalety poglądu russellovskiego. Wynikają z tego na przykład problemy epifenomenalizmu, których pogląd russellovski miał unikać wspomniane prawa wymagają też znacznego rozbudowania ontologii ponad i poza

być po prostu prawem to, że gdy stany mikrofenomenalne realizują stan informacyjny pewnego rodzaju na mocy związków przyczynowych między nimi (na mocy zasady „różnicy, która czyni różnicę”), powstaje bezpośrednia fenomenalna realizacja tego samego stanu. Rozwiązałoby to problemy teoretyczne kosztem skomplikowania ontologii. Nie mielibyśmy już prostej ontologii z fenomenologią będącą wewnętrznym aspektem fizycznie realizowanych informacji: część fenomenologii byłaby „dziwactwem” wyrosłym z tych informacji na mocy praw przypominających bardziej standardowy dualizm własności. Tak więc część zalet poglądu russellovskiego znikłaby, choć byłby on wciąż dość spójny.

W każdym razie pozostawiam to pytanie otwarte. Jest to z pewnością najtrudniejszy problem dla jakiegokolwiek typu poglądu russellovskiego; ale nie jest sprawą oczywistą, że nie można go rozwiązać. Gdyby tylko można było ją zastosować, druga strategia wydaje się szczególnie obiecująca; możliwe też, że do rozwiązania tego problemu potrzebny jest zupełnie nowy pomysł. Patrząc na sprawę optymistycznie, możemy postrzegać ten problem – uzgodnienia teorii psychofizycznej zarówno z faktami poziomu makro dotyczącymi naszego fenomenologii i jej podstaw fizycznych, jak i z poziomem mikro ontologii poglądu russellovskiego – jako jedno z podstawowych ograniczeń, które może doprowadzić nas do szczegółowej teorii świadomości. Jedną z trudności podczas tworzenia takiej teorii jest to, że nie ma zbyt wielu ograniczeń. Możliwe więc, że ten problem może dostarczyć tak potrzebnego zawężenia.

Jeśli żadna z tych strategii nie okaże się zadowalająca, będziemy musieli wycofać się z poglądu russellovskiego i przyjąć inne stanowisko metafizyczne. Można na przykład próbować opracować metafizykę czystej informacji jako sposób rozumienia świata fizycznego, a potem jakoś połączyć ją z fenomenologią, może na drodze opartego na prawach związku z czystą

własnościami, które są wymagane do ugruntowania fizyki. (Należy zauważyć, że Lockwood nie wykorzystuje tych praw do rozwiązania problemu ziarna; jego główną ideą rozwiązania tej kwestii jest intrygująca, odrębna propozycja z wykorzystaniem mechaniki kwantowej).

informacją. Można po prostu wycofać się do „oswojonej” ontologii zakładającej istnienie oddzielnej sfery fizycznej i fenomenalnej, z których każda ma własną wewnętrzną naturę, i między którymi istnieją związki oparte na prawach w duchu zasady informacji. Oznaczałoby to, że stwierdzenia o „podwójnym aspekcie” należałoby traktować mniej poważnie, a ontologia byłaby nieco mniej elegancka, ale mogłoby to nadal prowadzić do całkowicie zadowolającej teorii.

## 6 pytania otwarte

Naszkiecowane przeze mnie informacyjne ramy praw psychofizycznych pozostawiają ogromną liczbę otwartych pytań. Aby przekształcić tę wizję w teorię ostateczną, należałoby odpowiedzieć na wszystkie te pytania. O kilku problemach dotyczących ontologii tego stanowiska wspomniałem w poprzednim podrozdziale. Pojawia się też jednak wiele pytań dotyczących postaci *praw* oraz sposobu wyjaśniania naszej fenomenologii. Niektóre z nich to:

1. Dlaczego przestrzeń informacyjna realizowana fenomenalnie jest realizowana akurat tak, a nie inaczej? Na przykład, skoro nasza fenomenalna przestrzeń barw mogłaby być odwrócona, wydaje się nieco arbitralne, że jest taka, a nie inna. Czy trzeba dodać kolejne prawa, czy postulować przygodne „stałe”, aby ustalić te sprawy?

2. Czy charakter fenomenalnej przestrzeni informacyjnej jest wyznaczany przez strukturę przestrzeni (lub przynajmniej wyznaczany z dokładnością zapewniającą możliwość odwrócenia)? Mogłoby się wydawać, na przykład, że przestrzeń barw i przestrzeń smaków to proste trzy- lub czterowymiarowe przestrzenie, ale mają one bardzo różne charaktery, mimo podobnego kształtu. Można to uzasadniać w ten sposób, że podobieństwo struktur jest iluzją i że kiedy osadzi się obie w szerszej strukturze – postrze-

gając na przykład przeżycia barw jako składnik pełnej, głębokiej struktury przeżyć wzrokowych – to podobieństwo zniknie. Pozostaje jednak pytanie: Czy *jedyną* drogą rzutowania wzrokowych informacji barwnych na fenomenologię jest coś o charakterze zbliżonym do naszych przeżyć barwnych, czy też istnieje zupełnie inny sposób? Podejrzewam, że odpowiedź może być bliższa pierwszemu [członowi alternatywy], a nie drugiemu, ale nie jest wcale oczywiste, jak to uargumentować.

3. Za pomocą opisanych tu ram omawiałem przede wszystkim proste przeżycia postrzeniowe, takie jak przeżycia barw. Nie jest oczywiste, jak je rozbudować, aby uwzględniły subtelniejsze przeżycia, na przykład skomplikowane przeżycia emocjonalne i przeżycie bieżącej myśli. Czy można je tak rozbudować?

4. Jakiego rodzaju struktura formalna najlepiej ujmuje strukturę informacji fenomenalnych? Jakiego rodzaju przestrzenie topologiczne są potrzebne do uchwycenia relacyjnej struktury przeżyć? Czy powinniśmy przejść do struktury bardziej szczegółowej, np. przestrzeni metrycznej lub różniczkowej? Struktura kombinatoryczna przeżycia jest jeszcze bardziej interesująca: na przykład proste wielowymiarowe kontinuum jest prawdopodobnie ogromnym uproszczeniem struktury pola widzenia. Jak najlepiej uchwycić pełną strukturę? Czy w tym celu należy zmienić definicję przestrzeni informacyjnej?

5. Jak w tych ramach można ująć *jedność* świadomości? To znaczy, dlaczego moje przeżycia wzrokowe, słuchowe itd. są przeżyciami tego samego podmiotu? Podejrzewam, że w odpowiedzi należy uwzględnić sposób przetwarzania odpowiednich informacji, aby jedność świadomości odpowiadała faktowi, że odpowiednie informacje są dostępne do zintegrowania w określony sposób. Nie jest jednak jasne, jak miałoby się to sprawdzić w praktyce.

6. Jakie dokładnie są kryteria określające, które informacje w mózgu odpowiadają *moim* przeżyciom? Czy istnieje określona ścieżka przyczynowa lub szczególnego rodzaju strumień przyczynowy, które są tu istotne?

Prawdopodobnie coś w rodzaju bezpośredniej dostępności w globalnej kontroli odgrywa główną rolę w indywidualizacji informacji i odpowiednich ścieżek.

Istnienie wszystkich tych pytań wskazuje, jak daleko tym szkicowym pomysłom do prawdziwej teorii. Można to również dostrzec, zauważając, jak daleko tym ideom do możliwości przewidzenia na podstawie własności fizycznych systemu konkretnych własności fenomenalnych związanych z tym systemem fizycznym. W obecnej formie idea nie ma dużej *mocy* eksplanacyjnej i predykcyjnej: musi być znacznie wzmocniona, aby była naprawdę przydatna.

Potrzeba wielu nowych ustaleń, aby idea ta stała się zadowalającą teorią. Być może przełom mogłaby przynieść analiza problemu z poprzedniego podrozdziału: jak uzgodnić informacje fenomenalne w skali makroskopowej ze stanowiskiem informacyjnym [postulującym] „własności wewnętrzne” w skali mikroskopowej. Innego przełomu można by dokonać podczas poszukiwania ograniczenia, które daje klasę fizycznie zrealizowanych przestrzeni informacyjnych realizowanych w przeżyciu. Jeszcze inne przełomy mogą nastąpić w miejscach, których w ogóle nie wziąłem pod uwagę.

Idea może okazać się całkowicie błędna. Nie zdziwiłoby mnie to; faktycznie sądzę, że jest dosyć prawdopodobne, iż klucz do fundamentalnej teorii jest gdzie indziej. Zaproponowałem jednak te idee, gdyż musimy zacząć zastanawiać się nad tymi kwestiami, a także dlatego, że zobaczenie nawet nieudanego przykładu może być pouczające. Mam również nadzieję, że niektóre z zaproponowanych idei – dotyczących wyjaśniania sądów fenomenalnych, wszechobecności przeżyć, a także związku między przeżyciami, informacjami i własnościami wewnętrznymi tego, co fizyczne – mogą okazać się przydatne nawet po przeniesieniu do innych ram [teoretycznych]. Być może adekwatniejsza teoria świadomości mogłaby być utrzymana w duchu zbliżonym do przedstawionych tu pomysłów, nawet jeśli jej szczegóły będą zupełnie inne.

Często mówi się, że problem z teoriami świadomości tego typu polega na tym, że są one zbyt spekulatywne i niesprawdzalne. Sądzę jednak, że prawdziwy problem z zaproponowaną przeze mnie „teorią” jest inny: pozostaje ona zbyt ogólnikowa, gdy chodzi o przewidywanie. Gdybyśmy mieli teorię o porównywalnym poziomie prostoty, która mogłaby przewidzieć wszystkie konkretne fakty dotyczące naszych przeżyć – choćby tylko te znane z przypadku pierwszoosobowego – na podstawie faktów fizycznych dotyczących naszego systemu przetwarzania [informacji], byłyby to niezwykle osiągnięcie, dające nam dobry powód, by uznać tę teorię za prawdziwą. Obecnie nie dysponujemy taką teorią, ale nie należy sądzić, że nie da się jej stworzyć.