

Mirosław Zając

Przyczynowość stanów mentalnych

Nowa Krytyka 7, 55-72

1996

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Mirosław Zajac

Przyczynowość stanów mentalnych

Wydaje się, że termin „przyczynowość stanów mentalnych” powinien brzmieć dość dwuznacznie w ustach redukcjonisty. Jednak stosowanie go w różnych odmianach redukcjonizmu nie wynika tylko z archaicznych (dla niektórych) przyzwyczajęń. Pierwsza część tej pracy pokaże w skrócie, jakie motywy kierują redukcjonistą gdy używa on tego terminu i z jakimi problemami musi się uporać. Druga część będzie dotyczyła przyczynowości dwóch podstawowych własności stanów mentalnych (*qualia* i treści propozycjonalne), widzianej z perspektywy redukcjonizmu.

I

W 1981 r. Paul M. Churchland stwierdził, że kłopoty wiążące się z istnieniem tzw. *qualia*, dat zmysłowych, „surowych odczuć”, skończyły się dawno temu, co należy rozumieć, że samo ich istnienie jest wysoce problematyczne. Obecnym zagrożeniem dla teorii materialistycznych ma być dziedzina intencjonalności: postawy propozycjonalne¹. Wydawałoby się więc, że program redukcjonistyczny czy eliminatywistyczny powiódł się już przynajmniej co do tych pierwszych. Problem zatem zawęzałby się do natury zdań

¹ P.M. Churchland: *Eliminative Materialism and Propositional Attitudes*. „The Journal of Philosophy” 1981, vol. LXXVIII, s. 67.

będących sprawozdaniami ze swych własnych stanów psychicznych (m.in. do ich rzekomej nieobalalności). W takiej perspektywie, zagadnienie wzajemnej relacji między umysłem a ciałem ma charakter epistemologicznej dystynkcji na niepowątpiewalne sądy o stanach mentalnych i na – z zasady niepewne – sądy o stanach fizycznych. Takie stanowisko zajmuje Rorty. Dla niego, relacja ontologiczna między subiektywnymi a cielesnymi stanami opiera się na epistemologicznym charakterze obydwu stanów. Zdaniem Rorty'ego, problem (przestarzały) relacji umysł–ciało próbuje odpowiedzieć na pytanie, w jaki sposób pewne spostrzeżenie stanów subiektywnego świata jest połączone z niepewnymi spostrzeżeniami świata obiektywnego².

Można mieć zastrzeżenia wobec takiego opisu problemu umysł–ciało. Z historycznego punktu widzenia, najistotniejsze dla tego problemu jest poszukiwanie zrozumiałego związku pomiędzy niematerialnymi a materialnymi częściami pewnego *compositum* psychofizycznego. Z jednej strony, problem nie znika gdy odrzucimy kartezjańską tezę, iż konieczne jest pojmowanie tej niematerialnej części jako zdolnej do istnienia poza ciałem, lub że myślenie i doświadczenie zmysłowe jest wyłącznie z nią związane. Z drugiej strony, problem umysł–ciało nie sprowadza się jedynie do niepowątpiewalności zjawisk mentalnych. Tym, co pozostaje w tym problemie jako podstawa do epistemologicznej charakterystyki stanów mentalnych, jest postulat ich nieredukowalności do procesów materialnych.

Pozostawiając na boku względy adekwatnej interpretacji historycznej, trudno się dziwić stanowisku Rorty'ego. Po pierwsze: „interakcjonizm” nie jest dzisiaj cenionym terminem, jeśli jego konotacja dotyczy związku zdarzeń mentalnych ze zdarzeniami fizycznymi. Mało kto w obecnym czasie utrzymuje, że umysł i ciało są odrębnymi substancjami. Współczesne teorie tożsamości przyjmują, że stany mentalne muszą być identyczne ze stanami innego rodzaju, tak aby mieściły się w naturalistycznym poglądzie na świat. I aczkolwiek ten ostatni dopuszcza szerszy zakres

² R. Rorty: *Philosophy and the Mirror of Nature*. Princeton 1979, s. 54–59.

poglądów na relację między mentalnym i fizycznym niż tylko ścisła identyczność, m.in. funkcjonalizm (zazwyczaj związany z pewną formą identyczności), teorię pochodności, to wspólny im wszystkim jest redukcjonizm mentalnego do fizycznego³. Jednak we wszystkich powyższych teoriach (poza eliminatywizmem) pojęcie „przyczynowości mentalnej” znajduje zastosowanie⁴. A to rodzi przypuszczenie, że owo pojęcie odnosi się do zdarzeń, których moc sprawcza stoi pod znakiem zapytania, gdyż jak zrozumieć autentyczną przyczynowość zdarzeń mentalnych, skoro są one albo identyczne, albo zredukowane w inny sposób do zdarzeń fizycznych. Czy rzeczywiście chodzi nam o taką przyczynowość: zdarzenie mentalne M = zdarzenie neurofizjologiczne N, a więc zdarzenie M jest sprawcze, ponieważ N jest sprawcze?

Przyczyna powoduje skutek na mocy pewnych swoich własności. Przyczyną mojego nagłego osłepienia słońcem nie jest jego kulistość, lecz jaskrawość jego światła. Analogicznie, kiedy przyczyna mentalna powoduje jakiś skutek, to czy jest tak na mocy posiadania przez nią własności mentalnych czy niementalnych? Jest to pytanie o to, czy własności mentalne mogą być sprawcze. Mówiąc ściśle, w przypadku odrzucenia tej możliwości nie istnieje żadna prawdziwie mentalna przyczynowość, czyli nasze myśli i wrażenia nie wpływają przyczynowo na rzeczy.

Sprawę komplikuje fakt, iż pewne teorie identyczności, ich program redukcji, zakłada istnienie praw psychofizycznej korelacji, które stanowią istotny składnik wyjaśnienia psychofizycznej przyczynowości. Lecz postulowanie istnienia takich praw prowadzi do znanych trudności. Załóżmy, że zdarzenie mentalne typu M jest nomologicznie powiązane z jednoczesnym zdarzeniem neurofizjolo-

³ Teoria pochodności jako teza o niemożliwości niezależnej zmiany w dziedzinie mentalnej bez odpowiedniej zmiany w dziedzinie fizycznej nie wydaje się ewidentnie redukcjonistyczna. Jednak jest to teza ostrożnego redukcjonisty, jak zauważył D. Lewis, nie zaprzeczającego istnienia czegoś takiego jak stany mentalne, nie twierdzącego, iż istnieją one jedynie w pewien niefundamentalny sposób oraz nie parafrazującego określeń mentalnych poprzez terminy fizyczne. *New Work for a Theory of Universals*. „Australasian Journal of Philosophy” 1983, vol. 61, s. 358.

⁴ Dowodem tego może być m.in. wydany ostatnio zbiór rozpraw *Mental Causation*, ed. J. Heil. Oxford 1993.

gicznym N; założmy następnie, że M powoduje zdarzenie fizyczne (pewne zachowanie) N*. Przyjmując Hume'a objaśnienie przyczynowości, powinno istnieć prawo łączące M i N*. Jeżeli jednak istnieje związek nomologiczny między M i N, to z tego wynika, że istnieje także związek nomologiczny między N i N*. Powyższa sytuacja generuje trzy trudności.

(a) *Problem uprzedniości*. Czy przyczyną N* jest zarówno M jak i N? Ponieważ jednak istnieje identyfikowalna droga przyczynowa z N do N*, a co nie jest tak widoczne w przypadku M do N*, więc M wydaje się być wyprzedzona przez N, niejako „odgradza” ona M od N*⁵.

(b) *Problem nadokreślenia*. Zarówno M i N w powyższej sytuacji występują jako indywidualnie wystarczające dla wystąpienia N*. Jednak to wydaje się nieprawidłowe. Powodem mojego zachowania nie jest ból i neurofizjologiczne zdarzenie leżące u podstaw tego bólu.

(c) *Problem częściowej przyczynowości*. Skoro istnieje nomologiczny związek między M i N, nie wydaje się możliwe zaaranżowanie takiej sytuacji eksperymentalnej, gdzie każde z nich wystąpi osobno jako wystarczające dla spowodowania N*.

Jak widać, powyższe problemy są także wzajemnie sprzeczne. Przyczynowość mentalna staje pod znakiem zapytania⁶.

Pewnym przełomem są tu prace Davidsona (zwłaszcza „Zdarzenia mentalne”), które próbują pogodzić pozornie (według niego)

⁵ To samo rozumowanie stosuje się do własności funkcjonalnych lub dyspozycji. Jeżeli własność funkcjonalna „bycie lekiem przeciwbólowym” jest pochodna (drugiego rzędu) od własności strukturalnych np. aspiryny, to sprawczość tej struktury (uśmierzenie bólu) wydaje się „ogółać” ze sprawczości własność funkcjonalną.

⁶ J. Kim stwierdza, że owe problemy są do uniknięcia gdy teorię identyczności (w typie) zastąpimy poglądem o pochodności mentalnego od fizycznego. Trudno jednak zaakceptować jego stanowisko, kiedy sam określa je jako wersję epifenomenalizmu. Epifenomenalizm wzbrania się przed uznaniem przyczynowej roli stanów mentalnych. Tym bardziej nie pomagają zrozumieć jego stanowiska takie stwierdzenie, jak: przyczynowość zdarzeń mentalnych „zależy od przyczynowych władz leżących u podstaw procesów fizycznych” (48): *Causality, Identity, and Supervenience in the Mind-Body Problem*, [w:] *Midwest Studies in Philosophy*, IV. Minneapolis 1979, s. 31–49.

sprzeczne tezy. Jedna (*Zasada Interakcji Przyczynowej*) mówi, że „przynajmniej niektóre zdarzenia mentalne wchodzą w interakcje przyczynowe ze zdarzeniami fizycznymi”. Druga (*Zasada Nomologicznego Charakteru Przyczynowości*) stwierdza, że „tam, gdzie występuje przyczynowość, musi istnieć prawo”, a trzecia (*Zasada Anomalnego Charakteru tego, co Mentalne*) mówi, że „nie istnieją ściśle, deterministyczne prawa, na podstawie których można przewidywać i wyjaśniać zdarzenia mentalne”⁷. Sens tej „pozornej sprzeczności” jest następujący: jeżeli zachodzi interakcja przyczynowa pomiędzy zdarzeniami mentalnymi i zdarzeniami fizycznymi, to pewne zdarzenia mentalne są powodowane przez zdarzenia fizyczne. W takim przypadku muszą istnieć prawa leżące u podstaw tych związków przyczynowych. Wtedy jednak powinny istnieć prawa, na podstawie których można przewidywać i wyjaśniać zdarzenia mentalne, co jest sprzeczne z *Zasadą Anomalnego Charakteru tego, co Mentalne*.

Jego teoria anomalnego monizmu wydaje się usuwać napięcie, jakie istnieje między tymi zasadami (przez odrzucenie teorii identyczności typów i sceptycyzm wobec istnienia praw rządzących zdarzeniami mentalnymi). Lecz dzieje się tak za pewną cenę. Przedstawiając rzecz ściślej, teoria anomalnego monizmu zawiera trzy tezy: każde niepowtarzalne zdarzenie mentalne jest identyczne co do egzemplarza z pewnym zdarzeniem fizycznym; następnie, każde pojedyncze stwierdzenie przyczynowe jest *ekstensjonalne*, tzn. denotuje relację zachodzącą pomiędzy zdarzeniami jednostkowymi, niezależnie od tego, jak się je opisuje; i w końcu, zdania zawierające predykaty wiążące zdarzenia z prawami są *intensjonalne*, „zdarzenia mogą więc stanowić przypadki praw, i tym samym mogą być wyjaśniane i przewidywane w ich świetle tylko wtedy, gdy są one opisane w taki lub inny sposób”⁸.

Ostatnia teza wyraźnie wskazuje, że wbrew temu, co wydaje się sugerować *Zasada Anomalnego Charakteru tego, co Mentalne*, zdarzenia mentalne nie posiadają anomalii „w sobie”, nie występują

⁷ D. Davidson: *Zdarzenia mentalne*, [w:] *Eseje o prawdzie, języku i umyśle*. Warszawa 1992, s. 164–166.

⁸ D. Davidson: *op.cit.*, s. 176.

poza zakresem praw naturalnych. W przeciwnym razie konflikt między tymi zasadami byłby nierozwiązywalny. Aby móc wyjaśnić i przewidywać pewne zdarzenie na podstawie praw, zdarzenie to musi wystąpić pod odpowiednim opisem, najlepiej gdy opis wykorzystuje słownik praw przywoływanych dla objaśnienia lub predykcji. Zatem „anomalność” tego, o którym mówi powyższa zasada, nie może dotyczyć samego zdarzenia mentalnego, lecz raczej jego mentalistycznego opisu. Dlatego L. Antony proponuje zastąpić ową zasadę lub przynajmniej tak ją rozumieć, że implikuje ona następującą zasadę: „Nie istnieją ściśle prawa deterministyczne formułowalne w słownictwie psychologicznym”⁹. Gdyby istniały takie prawa, zdarzenia opisane mentalistycznie nie byłyby w żadnym sensie „anomalne”.

Ostatecznie więc jądro anomalnego monizmu można przedstawić w taki sposób. Prawdziwość pojedynczego stwierdzenia przyczynowego wymaga istnienia prawa naturalnego łączącego dwa zdarzenia. Nie ma żadnego (co wynika z powyższej poprawionej zasady) prawa formułowalnego w terminach psychologicznych, lecz ponieważ każde zdarzenie mentalne jest identyczne z pewnym zdarzeniem fizycznym, każda psychofizyczna interakcja jest identyczna z pewną podpadającą pod prawo fizyczną interakcją.

Davidson widzi swoją teorię jako reakcję na radykalny redukcjonizm (sam określa ją jako „łagodny monizm”), teorię pozwalającą pogodzić możliwość materializmu z nieredukowalnością stanów mentalnych. Jednak – co słusznie podkreśla Antony – błędem byłoby sytuować teorię Davidsona w pobliżu funkcjonalizmu, gdyż odrzuca on coś więcej niż tylko radykalny redukcjonizm i to z powodów, na które nie przystałby funkcjonalista¹⁰. To jest właśnie ta cena, którą płaci Davidson w zamian za możliwość używania terminu „przyczynowość mentalna” w sposób (relatywnie) niekontrowersyjny. Davidson mianowicie odrzuca prawa redukcyjne na

⁹ L. Antony: *Anomalous Monism and the Problem of Explanatory Force*. „The Philosophical Review” 1989, vol. XCVIII, s. 163. Antony przekonująco dowodzi, że chociaż anomalny monizm jest niesprzeczną teorią, to stoi on w sprzeczności z przyjmowaną przez Davidsona przyczynową teorią działań.

¹⁰ L. Antony: *op.cit.*, s. 174–183.

podstawie uprzedniego odrzucenia psychologii *tout court*. Argumentem dla funkcjonalisty za odrzuceniem radykalnego redukcjonizmu jest możliwość (czy prawdopodobieństwo) wielorakiej (fizycznej) realizacji stanów mentalnych, natomiast dla Davidsona jest to niemożliwość istnienia jakichkolwiek praw psychologicznych. Ta ostatnia wynika z faktu, iż zdarzenia mentalne nie tworzą „systemu zamkniętego” i tu nie chodzi tylko o to, że wchodzą one w interakcje ze zdarzeniami niementalnymi, lecz o bardziej fundamentalny – według Davidsona – fakt, że istnieją zdarzenia, do opisu których język psychologii jest nieadekwatny¹¹. Davidson przyjmuje, że fizyka nie podlega tym ograniczeniom, tj. jakkolwiek proces w naturze poddaje się fizykalistycznej charakterystyce.

Ważne jest dostrzeżenie, że chociaż fizyka w powyższym sensie jest dla Davidsona nauką podstawową, to główny morał anomalnego monizmu odnośnie psychologii zawiera się w stwierdzeniu, iż niemożliwa jest jej redukcja do danej gałęzi nauk przyrodniczych. Jednak jeżeli teoria dat zmysłowych jest integralną częścią psychologii, to także *qualia* bronią się przed redukcją, o czym więcej za chwilę.

Bez względu na to, jakie motywy kierują redukcjonistą w stosowaniu terminu „przyczynowość mentalna” (od zdroworozsądkowego poczucia, że to nasze stany umysłowe powodują nasze zachowanie, przez stosowanie tego terminu dla celów czysto redukcjonistycznych, aż do funkcjonalistycznej konstatacji, że rzeczywistość stanów mentalnych wyczerpuje się wyłącznie na ich roli przyczynowej), faktem pozostaje, że ów termin zachował swoją ważność, chociaż może już nie w swej kartezyjańskiej konotacji. Dalsza część tego artykułu odda ten stan rzeczy przez analizę tego, w jaki sposób redukcjoniści pojmują przyczynowość dwóch podstawowych własności stanów mentalnych: dat zmysłowych i treści. Testem dla nich będzie to, w jaki sposób radzą sobie z naszym subiektywnym, lecz mocnym odczuciem, że nasze stany mentalne są tym, czym są

¹¹ Zob. D. Davidson: *op.cit.*, s. 190–191.

i nie mogłyby być inne, można by rzec: redukcjoniszc vs. „prywatny esencjalizm”.

II

Przyczynowość dat zmysłowych. Czy rzeczywiście kłopoty z istnieniem dat zmysłowych są już nieaktualne, jak zadekretował jakiś czas temu Churchland? W tym miejscu potrzebny jest przegląd tego, w jaki sposób opierały się one zapędowi redukcjonistycznym.

Qualia są fenomenalnymi własnościami wrażeń i stanów percepcji. Najlepszym przykładem takich stanów są te, które zwykle opisujemy za pomocą jakościowych, fenomenalnych cech: doświadczenie zmysłowe, wyobrażenie mentalne. Nasze odczucie, że przynajmniej pewne stany mentalne są autentycznie *wewnętrznymi* stanami, do których mamy dostęp w introspekcji, było głównym powodem odejścia od behawioryzmu w stronę teorii identyczności typów. Ta ostatnia mówi coś takiego: pewien typ stanów mentalnych, np. ból, jest identyczny z pewnym typem stanów fizycznych, np. pobudzeniem C-włókien nerwowych. Ponieważ zarówno ból, jak i stymulacja C-włókien nerwowych są rodzajami naturalnymi, identyczność pomiędzy tymi rodzajami pociąga za sobą to, że każdy egzemplarz jednego rodzaju musi być w tym samym czasie egzemplarzem drugiego rodzaju. Z tego wynikają dwie dalsze konsekwencje, po pierwsze, nie może istnieć ból, który nie jest pobudzeniem C-włókien nerwowych, i po drugie, nie mogą istnieć pobudzenia C-włókien nerwowych, które nie są bólem.

Przypomnijmy dwa główne zarzuty jakie postawiono tej teorii. Pierwszy opiera się na powszechnie uznanym fakcie, iż przynajmniej pewne typy stanów czy własności mentalnych dopuszczają wieloraką realizowalność, jeden i ten sam stan mentalny może być realizowany na różne fizyczne sposoby. Jeżeli ośmiornica lub jakakolwiek inna wyobrażalna istota (np. marsjanin) odczuwa ból czy głód, powinna ona – w myśl tej teorii – posiadać te same stany

fizykochemiczne (mózgu) co człowiek, a to jest nieprawdopodobne¹². Podobną krytykę, wykorzystującą teorię bezpośredniej referencji, znajdziemy u Kripkego. Wykorzystuje on kartezjańska intuicję, że ból mógłby istnieć pod nieobecność zjawiska fizycznego. Zarzut Kripkego można oddać w takim oto ciągu rozumowań. Jeżeli dwa sztywne desygnatory desygnują ten sam rodzaj naturalny w tym świecie, to desygnują one ten sam rodzaj naturalny w każdym możliwym świecie, w którym ten rodzaj istnieje; zarówno „ból” jak i „pobudzenie C-włókien nerwowych” są sztywnymi desygnatorami rodzajów naturalnych. Możemy wyobrazić sobie świat, w którym jest ból, lecz żadnego pobudzenia C-włókien nerwowych lub odwrotnie, wówczas „ból” i „pobudzenie C-włókien nerwowych” nie desygnują tego samego rodzaju naturalnego w każdym możliwym świecie, w którym ów rodzaj istnieje. Stąd „ból” i „pobudzenie C-włókien nerwowych” nie desygnują tego samego rodzaju naturalnego, a więc materialistyczne stwierdzenie identity ciała i umysłu jest fałszem¹³. Zarzut Kripkego, jeśli jest adekwatny, dowodzi jednak czegoś, czego nawet umiarkowany redukcjonista nie może przyjąć, a mianowicie słuszności kartezjańskiego dualizmu.

Drugi typ zarzutu, tzw. „argument z wiedzy”, sprowadza się do tezy, że żadna ilość informacji fizycznej nie daje nam wiedzy o jakościowych cechach wrażeń, zatem własności fenomenalne nie są własnościami fizycznymi. Najbardziej znany eksperyment myślowy Jacksona zakłada, że przy pewnych warunkach (zamknięcie w czarno-białym pomieszczeniu, kontakt ze światem za pomocą czarno-białego monitora itp.) dana osoba może posiadać wiedzę o wszystkich własnościach fizycznych percepcji wizualnej, włączając w to wiedzę o zdarzeniach zachodzących w mózgu osoby spostrzegającej. Teraz pojawia się pytanie: czy ta osoba uwolniona z tego pomieszczenia dowie się czegoś nowego? Jeśli tak, to przede wszystkim dowie się czegoś o kolorach, czyli o własnościach wrażeń. Ponieważ jednak jej poprzednia wiedza dotyczyła, co

¹² Zob. np. H. Putnam: *The Nature of Mental States*, [w:] *Readings in Philosophy of Psychology*, ed. N. Block. Vol. I. Cambridge, Mass 1980, s. 223–231.

¹³ Zob. S. Kripke: *Naming and Necessity*. Cambridge, Mass. 1980, s. 148–154.

zakładała hipoteza, wszystkich, istotnych dla tego eksperymentu własności fizycznych, *nowe* własności nie mogą być natury fizycznej¹⁴.

Ten zarzut spotkał się z różnego rodzaju sprzeciwami. Twierdzono na przykład, że wiedza o pewnej danej zmysłowej (np. jak smakuje dana potrawa) nie jest w ogóle sprawą posiadania informacji, lecz raczej posiadania pewnych zdolności (dyspozycji). Twierdzono także, że posiadanie nowej wiedzy o czymś nie jest tym samym, co posiadanie wiedzy o czymś nowym. Wszystkie te obiekcje muszą być jednak ostrożnie rozważane. Przypomnijmy, że *quale* jest *fenomenalnym aspektem* wrażeń, czyli jest to własność wrażeń, które są w pewien sposób odczuwane czy doświadczane (fenomenalnie, subiektywnie). Nasze doświadczanie pewnego *quale* jest samą jego istotą, jego *esse* jest *percipi*. Zdolność do wyobrażania, rozpoznawania, przewidywania, które miałyby się wiązać z wiedzą o danych zmysłowych zakłada jednak dalsze stany, które także są w pewien sposób odczuwane. Co do drugiego sprzeciwu, konkluzja Jacksona nie jest natury epistemologicznej czy konceptualnej, lecz ontologicznej: do samej natury danych zmysłowych należy to, że są one spostrzegane w pewien sposób¹⁵.

Przyjmuje się, że daty zmysłowe stanowią jeszcze większy problem dla funkcjonalizmu, niż dla fizykalistycznej teorii identyczności typów. Funkcjonalizm w filozofii umysłu utrzymuje, że stany mentalne są stanami funkcjonalnymi, stanami, które są charakteryzowane przez ich miejsce w związkach przyczynowych, a więc stanem mentalnym jest cokolwiek, które pełni odpowiednią rolę przyczynową. Warunek wielorakiej realizowalności – sam motor funkcjonalizmu – wydaje się jeszcze do pogodzenia z takimi intencjonalnymi stanami jak przeświadczenia czy pragnienia, lecz nie jest to już takie oczywiste w przypadku *quale*. *Qualia* są tym, czym

¹⁴ F. Jackson: *Epiphenomenal Qualia*. „Philosophical Quarterly” 1982, vol. 32, s. 130.

¹⁵ Dyskusję za i przeciw tego typu zarzutom znajdziemy np. [w:] C. Peacocke, recenzja z D. Lewisa *Philosophical Papers*. „Journal of Philosophy” 1985, vol. 82, s. 42–45; A. Marras: *Materialism, Functionalism and Supervenient Qualia*. „Dialogue” 1993, vol. XXXII, s. 475–492.

są dzięki „monadycznej” (wewnętrznej) naturze wrażeń, których są własnościami. Funkcjonalizm natomiast dostrzega tylko te stany, których własności są zewnętrzne, relacyjne. Mamy raczej mocną intuicję, że różne rodzaje biologicznych istot nie muszą dzielić z nami takich samych dat zmysłowych. Co więcej, można by twierdzić, że mogłyby istnieć systemy funkcjonalnie nieodróżnialne od nas samych, gdzie przypuszczenie, iż w ogóle doświadczają one pewnych dat zmysłowych (nie mówiąc już o takich samych jak nasze) jest wielce nieprawdopodobne¹⁶.

Z powyższego wynika, że czysto funkcjonalna charakterystyka stanów mentalnych nie potrafi zająć się ich podstawowym aspektem, własnościami fenomenalnymi. W myśl tej teorii, ból okaże się jakimkolwiek stanem, który pełni przyczynową rolę, określona przez potoczne pojęcie bólu. W efekcie: cokolwiek okaże się bólem będzie czymś, które może nie być bólem. Z subiektywnej perspektywy, opis bólu jako stanu, który faktycznie mógłby nie być bólem jest nie do zaakceptowania. Ponadto, jeżeli pojęcia stanów mentalnych są czysto funkcjonalne, to musimy utożsamić te stany z pewnymi innymi stanami, ponieważ czysto funkcjonalne stany są abstrakcyjne i stąd niezdolne do pełnienia takich ról.

Redukcja stanów mentalnych w funkcjonalizmie zazwyczaj przybiera formę identyczności w egzemplarzu, tzn. zdarzenia mentalne (włącznie z ich aspektami fenomenalnymi) są identyczne co do konkretnych przypadków ze zdarzeniami fizycznymi. Wówczas jednak pojawia się zagrożenie epifenomenalizmu, zdarzenia mentalne nie odznaczają się sprawczością przyczynową. Rozpatrzmy taką sytuację: zakładamy, że pewne zdarzenie mentalne z_1 (typu M (odczucie bólu) jest przyczyną pewnego zachowania z_2 (wycofanie ręki). Załóżmy dalej, że z_1 jest identyczne w egzemplarzu ze zdarzeniem neurofizjologicznym typu N. Okazuje się, że z_1 spowodowałoby z_2 nawet wtedy, gdyby z_1 nie było wrażeniem bólu (tj. egzemplarzem M), oczywiście tak długo, jak długo było

¹⁶ N. Block wymyślił system funkcjonalny, który miał być realizowany przez ludność Chin w taki sposób, aby kolektywne zachowanie naśladowało funkcje mentalne. *Troubles with Functionalism*, [w:] *Readings in Philosophy...*

zdarzeniem neurofizjologicznym typu N. Stąd dla z_1 powodującego z_2 , bycie egzemplarzem M nie jest wcale istotne, jeżeli $M \neq N$.

Wydawałoby się więc, że fizykalizm identyczności typów lepiej sobie radzi z przyczynowością stanów mentalnych (wystarczy zgodzić się na $M = N$), gdyby nie wątpliwości związane z tą teorią (patrz część I tej pracy).

Czy jednak mentalista, czyli ktoś utrzymujący, że stany mentalne są *sui generis* (oraz że cechuje je sprawczość przyczynowa), sam nie wpada w pułapkę epifenomenalizmu? Taki jest wniosek przejrzystego rozumowania T. Horgana¹⁷. Według niego, jeżeli zanegujemy, że stany mentalne są albo identyczne, albo pochodne od innych stanów, to nie mogą one spełniać żadnej roli przyczynowej, a ściślej: pochodność dat zmysłowych jest koniecznym warunkiem ich sprawczości przyczynowej, stąd są one pochodne. Argumentacja Horgana zakłada, że mechanizm jest prawdą, gdzie przez „mechanizm” rozumie on, z grubsza biorąc, determinizm fizyczny. Teza o pochodności dat zmysłowych brzmi następująco:

(PQ) Nie istnieją dwa fizycznie możliwe światy, które są dokładnie podobne pod względem fizykochemicznym, lecz odmienne co do występowania dat zmysłowych.

Przedstawmy samą istotę tego rozumowania. Jeżeli *qualia* nie są pochodne od stanów fizycznych, korelacja pomiędzy *qualia* a stanami fizycznymi będzie się różniła od jednego możliwego świata do drugiego. Będą możliwe takie światy, gdzie jedyną różnicą jest różnica pomiędzy nimi, czyli będą możliwe dwa światy, które różnią się radykalnie w ich korelacji *qualia*/stany fizyczne, lecz pod innymi względami są dokładnie podobne. W takim przypadku jednak *qualia* nie mogłyby mieć żadnej roli przyczynowej, gdyż w obydwu światach będą obowiązywały te same prawa przyczynowe, pomimo zmian w występowaniu *qualia*.

¹⁷ T. Horgan: *Supervenient Qualia*. „The Philosophical Review” 1987, vol. XCVI, s. 491–520.

Niewątpliwie sam mechanizm, co przyznaje Horgan, generuje znany problem, iż *qualia* mogłyby nie posiadać żadnej roli przyczynowej na tle przyczynowości leżących u ich podstaw zdarzeń neurofizjologicznych. Pogodzenie mechanizmu ze sprawczością dat zmysłowych skłania do przyjęcia teorii identyczności psychofizycznej (albo w jej wersji identyczności w typie lub przynajmniej identyczności w egzemplarzu). Co więcej, Horgan dowodzi, że pochodność jest koniecznym, lecz nie wystarczającym warunkiem przyczynowości dat zmysłowych. Którą wersję teorii identyczności należy tu zastosować, pozostawia kwestią otwartą.

Przyczynowość treści propozycjonalnych. Problem przyczynowości treści stanów mentalnych okazuje się jeszcze bardziej skomplikowany niż w przypadku dat zmysłowych. Zaczniemy od naszych potocznych odczuć. Tak jak wydaje się dla nas jasne, iż fenomenalny aspekt naszych doświadczeń ma wpływ na nasze zachowanie, podobnie jest dla nas niewątpliwe, że np. zamiar podniesienia mojej lewej ręki jest przyczyną intencjonalnego podniesienia tej ręki przeze mnie. Natychmiast jednak pojawiają się różnego rodzaju problemy.

Jedne dotyczą zagadnienia, w jaki sposób „semantyczne” własności stanów mentalnych są odpowiedzialne za skutki powodowane przez te stany. Nie jest łatwo zrozumieć, jak stany, które reprezentują świat mogą powodować skutki przez to, że reprezentują świat. Stany intencjonalne są relacjami do sądów (ich charakter jest z istoty relacyjny) i na mocy tych relacji reprezentują świat. Lecz cokolwiek uważamy za sądy – myśli Fregego, sądy Russella czy zbiory możliwych światów – pozostają one przedmiotami abstrakcyjnymi bez parametrów czasoprzestrzennych. W jaki sposób relacja do przedmiotów abstrakcyjnych może powodować jakiegokolwiek skutki? Tak więc, bezpiecznym wyjściem jest stwierdzenie, że stany intencjonalne nie powodują skutków *na mocy ich treści*, a treści są tylko istotne dla wyjaśniania przyczynowego.

Jednak uzasadnienie powyższego stwierdzenia przez przyjęcie pewnej wersji teorii identyczności napotyka znane problemy. Co więcej, eksternalizm chce pozbawić redukcjonistę tej resztki złu-

dzeń, którą żywił on względem teorii pochodności. Jeżeli ekstermalizm jest właściwym poglądem na indywidualizację treści postaw propozycjonalnych, to same treści są wysoce relacyjne: dwa fizycznie nieodróżnialne podmioty mogą mieć odmienne treści przekonań, gdy treści te odnoszą się do odmiennego środowiska fizycznego lub są związane z różnym środowiskiem lingwistycznym¹⁸. Skoro przyjmuje się, że duplikaty mają przeświadczenia z różnymi treściami, a zarazem ich stany mentalne pozostają przyczynowe, to z tego wynika (dla niektórych), że nauki kognitywne nie powinny odróżniać stanów mentalnych tych duplikatów. Zatem różnice w treściach nie powodują żadnych różnic we własnościach przyczynowych stanów mentalnych, a więc treści nie charakteryzują się sprawczością przyczynową¹⁹.

Chociaż przyczynowość treści stanów mentalnych negowana jest dość powszechnie, to jednak istnieją wyjątki. Na przykład F. Jackson znajduje niezadowolającą taką sytuację, gdzie treść odgrywa pewną rolę w wyjaśnianiu zachowania, lecz sama jest pozbawiona sprawczości przyczynowej²⁰. Stanowisko Jacksona obejmuje funkcjonalizm i teorię identyczności w typie. Funkcjonalizm jest rozumiany w szerszy sposób niż zazwyczaj, tak aby nie był sprzeczny z zasadniczymi tezami ekstermalizmu (jednak zachowuje tzw. *narrow content*, która jest taka sama u duplikatów). Jackson utrzymuje cztery podstawowe sądy: własności są sprawcze; własności funkcjonalne nie są sprawcze; rola funkcjonalna wewnętrznych stanów podmiotu określa ich treści; treści są sprawcze. Pozorna sprzeczność między tymi sądami wynika, według niego, z nierozpoznania faktu, iż bycie zdeterminowanym przez rolę funkcjonalną nie jest tym samym, co bycie własnością funkcjonalną. Bycie zdeterminowanym przez rolę funkcjonalną nie stoi w sprzeczności z utożsamieniem własności intencjonalnych treści z własnościami neurofizjologicznymi, a te ostatnie są sprawcze.

¹⁸ Zob. T. Burge: *Individualism and the Mental*, [w:] *Midwest Studies in Philosophy*, IV. Minneapolis 1979, s. 73–121.

¹⁹ Zob. np. J. Fodor: *Psychosemantics*. Cambridge, Mass. 1987, s. 33–35.

²⁰ F. Jackson: *Essentialism, Mental Properties and Causation*. „Proceedings of the Aristotelian Society” 1995, vol. XCV, s. 253–268.

Należy się wyjaśnienie, dlaczego własności funkcjonalne, zdaniem Jacksona, nie mogą być sprawcze. Ma to być niezgodne z tezą Hume'a, że związki przyczynowe są przygodne, tj. własności posiadają moc sprawczą jedynie akcydentalnie. Dla każdego stanu jest prawdą, że ten sam stan mógłby nie spowodować tego, co faktycznie spowodował. „Tak więc, jeżeli pewna struktura chemiczna jest silnie trująca, zatem może być prawidłowo określona jako «zabójcza struktura», stwierdzenie «zabójcza struktura zwykle zabija» jest z konieczności prawdziwe. Jednak nie jest własnością istotną tej struktury, że zwykle zabija po jej spożyciu. Jest to własność akcydentalna. Istnieją możliwe światy, gdzie ta sama struktura jest nieszkodliwa” i gdzie samo jej określenie jako „zabójczej struktury” jest oczywiście niewłaściwe²¹.

Teraz Jackson musi się zmierzyć ze standardowym zarzutem podnoszonym przeciwko teorii identyczności w typie. W myśl tej obiekcji, możliwość wielorakiej realizacji wskazuje, że skoro mógłbym żywić przeświadczenie, że śnieg jest biały, podczas gdy nie istnieje odpowiednie pobudzenie moich B-włókien nerwowych, to pozostajemy z zagadką, jak to jest możliwe, że to przeświadczenie mogłoby być inne niż faktycznie jest. Według Jacksona jednak, zagadka powstaje wtedy, gdy teoria identyczności w typie jest interpretowana przez esencjalistę. Teoria ta mówi tyle, że własność danego stanu, iż jest on moim przekonaniem, że śnieg jest biały, jest własnością akcydentalną tego stanu. Możliwość, że pewien inny stan mógłby spełniać rolę przekonania, że śnieg jest biały, jest po prostu możliwością, że „moje przekonanie, że śnieg jest biały” mogłoby desygnować pewien inny stan, aniżeli ten, który obecnie desygnuje. Zatem prawdziwość identyczności:

Moje (oraz w X, Y,...) przekonanie, że śnieg jest biały =
pobudzenie B-włókien nerwowych

jest przypadkowa i *a posteriori*. Jackson podkreśla, że nie znaczy to, iż identyczność danego przekonania z samym sobą jest przypadkowa i *a posteriori*, niemniej jednak wskazuje, że teoria identyczności nie jest nazwana najszcześliwiej. Oczywiście dla wielu myśl,

²¹ Ibidem, s. 257.

że ich przekonanie, iż śnieg jest biały nie jest z *istoty* ich przekonaniem, że śnieg jest biały, może się wydawać nazbyt ekstrawagancka. Jackson jest jednak niewzruszony w swej pewności, iż nieprzyjęcie tego stanowiska gwałci tezę Hume'a, że to, co pewien stan czyni, nie jest jego istotną cechą.

Strategia Jacksona jest więc następująca. Z pozycji funkcjonalizmu i teorii identyczności w typie próbuje dowieść, że treści są przyczynowe. Ale, jak pamiętamy, funkcjonalizm w filozofii umysłu pojawił się właśnie jako reakcja na trudności trapiące identyfikację w typie, w szczególności problem wielorakiej realizowalności stanów mentalnych. Czyli musi on, z jednej strony, charakteryzować stany mentalne funkcjonalistycznie, a z drugiej strony, starać się zachować tezę o identyczności w typie. Gdyby jednak treści traktować jako własności funkcjonalne, taka identyfikacja nie byłaby możliwa, ponieważ własności funkcjonalne nie są, zdaniem Jacksona, przyczynowe. Stąd, treści będąc jedynie zdeterminowane przez rolę funkcjonalną wewnętrznych stanów podmiotu (rola ta ma być jednak tak rozumiana, aby czynić zadość wymogom eksternalizmu) mogą pozostać przyczynowe i zarazem podlegać wielorakiej realizacji.

Mam dwie uwagi odnośnie powyższego rozumowania. Pierwsza dotyczy wyjaśniania związku przyczynowego. Załóżmy, że czujnik elektroniczny brzęczy, gdy jego komórka fotoelektryczna napotka żółty przedmiot. Przyjmujemy także, że skład chemiczny tego przedmiotu jest wystarczający dla dania mu żółtego koloru. Czy zgodzimy się na prawdziwość takiego zdania: „Skład chemiczny tego żółtego przedmiotu jest tym, co powoduje brzęczenie czujnika”? Raczej nie, gdyż czujnik nie jest „wrażliwy” na skład chemiczny żółtego przedmiotu. *Kolor* przedmiotu jest tym, co powoduje reakcję czujnika. W tym przypadku zatem, to nie baza kategorialna (struktura chemiczna) jest przyczyną, lecz jego „funkcjonalny ekwiwalent”.

Druga uwaga podnosi znacznie poważniejszą kwestię. Skoro stwierdzenie:

Moje (oraz w X, Y,...) przekonanie, że śnieg jest biały =
pobudzenie B-włókien nerwowych

jest prawdziwe *a posteriori*, to nie jest ono prawdą konceptualną (gdyż te ostatnie są poznawalne *a priori*). Przyjmijmy, że wyrażenie ma *znaczenie niezależne od kontekstu*, gdy znaczenie tego wyrażenia nie zmienia się z kontekstem wypowiedzi, tj. ma jedno i to samo znaczenie we wszystkich (aktualnych i możliwych) kontekstach wypowiedzi. Można traktować jako zasadę, że prawdziwe zdanie ze znaczeniem niezależnym od kontekstu jest poznawalne *a priori*. Wobec tego, powyższa identyczność Jacksona jest zależna od kontekstu. Jak dotąd, Jackson nie powinien wnosić zastrzeżeń do tego typu rozumowania. Wydaje się oczywiste, że jeżeli zdanie jest zależne od kontekstu, to jest tak, ponieważ pewne jego części są zależne od kontekstu. W identyczności Jacksona mamy wybór, albo „przekonanie, że śnieg jest biały”, albo „pobudzenie B-włókien nerwowych” jest zależne od kontekstu.

Dość powszechnie przyjmuje się, że takie wyrażenia mentalistyczne jak: „zły na (coś)”, „obawia się”, „boli” mają znaczenia niezależne od kontekstu. Gdybyśmy zastąpili pierwszy człon identyczności Jacksona wyrażeniem „ból”, to trudno byłoby sobie wyobrazić, że w pewnym kontekście, ktoś mógłby używać wyrażenia „ból” z jego zwykłym znaczeniem, lecz odnosząc się do czegoś innego niż ból. Podobnie nie można sobie wyobrazić, że nasz fenomenologiczny bliźniak mógłby to robić odnosząc się do czegoś innego. W tym sensie, „ból” spełnia także wymogi bycia sztywnym desygnatorem. Nie jest to takie jasne, *prima facie*, w przypadku postaw propozycjonalnych. Lecz jeżeli zastosujemy podobny test jak w przypadku „ból” do samych postaw propozycjonalnych (bo przecież nie do sądów, z którymi są związane), to uzyskamy ten sam rezultat. Wyobraźmy sobie możliwy kontekst, w którym ktoś wypowiada słowa „moje przekonanie, że...” z aktualnym znaczeniem, lecz nie denotując swego przekonania. Przyjmując powyższe, ktoś mógłby wypowiedzieć „moje przekonanie” mając na myśli to, co my mamy, lecz nie stwierdzając, że ktoś ma przekonanie. To wydaje się absurdalne. A więc wyrażenie „moje przekonanie” jest niezależne od kontekstu.

O nazwach indywidualnych, które są niezależne od kontekstu i są sztywnymi desygnatorami można powiedzieć, że odnoszą się

do czegoś w sposób istotny²². Jest to możliwe, ponieważ istnieje szczególnie bliski związek pomiędzy znaczeniem nazwy a jej referentem; to znaczenie niejako wymusza, żeby referentem nie było nic innego, jak ta właśnie rzecz. Tak też jest w przypadku czyjegós przekonania, czyli nie jest prawdą stwierdzenie, że czyjeś przekonanie, że śnieg jest biały nie jest z *istoty* jego przekonaniem, że śnieg jest biały.

²² Por. E. Hirsch: *Metaphysical Necessity and Conceptual Truth*, [w:] *Midwest Studies in Philosophy*, XI. Minneapolis 1986, s. 249; R. Stalnaker dotyka tego problemu analizując pojęcie tzw. *narrow content* (w odróżnieniu od *wide content*, która jest idywidualizowana eksternalistycznie), dochodząc jednak do odmiennego wniosku. Według niego wewnętrzna nie jest tylko sama treść, lecz własność posiadania myśli z tą treścią; *Narrow Content*, [w:] *Propositional Attitudes: The Role of Content in Logic, Language and Mind*, eds. C.A. Anderson i J. Owens. Stanford 1990, s. 131–145.