Witold Hensel

Dwa funkcjonalizmy Hilary'ego Putmana, czyli kawałek historii z morałem

Diametros nr 29, 31-49

2011

Diametros № 29 (September 2011): 31-49

DWA FUNKCJONALIZMY HILARY'EGO PUTNAMA, CZYLI KAWAŁEK HISTORII Z MORAŁEM

– Witold M. Hensel –

1. WSTĘP

Kariera funkcjonalizmu w filozofii umysłu była szybka i błyskotliwa: już w kilka lat po opublikowaniu *Psychological Predicates* (Putnam [1967]) niemal zgodnie uznano, że teorię identyczności (inaczej fizykalizm typów) wraz z behawioryzmem logicznym należy porzucić. Na placu boju pozostał tylko funkcjonalizm, który dziś stanowi szeroki i zróżnicowany nurt filozoficzny. Ten natychmiastowy sukces zainteresowałby socjologów, lecz i filozofowi powinien dać do myślenia. Czy można racjonalnie przyjąć jakikolwiek pogląd w równie ekspresowym tempie?

Celem niniejszego artykułu[1] jest analiza i ocena argumentacji prowadzącej Hilary'ego Putnama (a za nim innych) do przyjęcia funkcjonalizmu. Zaczynam od naszkicowania sytuacji problemowej przed pojawieniem się tej koncepcji (część 2.); w części 3. przedstawiam wczesne stanowisko Putnama, które nazywam „słabym funkcjonalizmem". Był to pogląd sceptyczny: z jednej strony wykorzystywał intuicję, że stany umysłu mają charakter relacyjny, do obalenia behawioryzmu logicznego i fizykalizmu typów[2], z drugiej zaś wskazywał na arbitralność wszelkich rozstrzygnięć w kwestii umysł–ciało. Słaby funkcjonalizm był, zdaniem Putnama, stanowiskiem niekonsekwentnym, dlatego w naturalny sposób przekształcił się w doktrynę pozytywną, odrzucającą tezę o arbitralności rozwiązań w filozofii umysłu.

[1] Chciałbym gorąco podziękować Joannie Mach-Komorowskiej, Marcinowi Jażyńskiemu, Marcinowi Miłkowskiemu i anonimowemu recenzentowi ICF „Diametros" za lekturę wcześniejszej (w wypadku Marcina Miłkowskiego: wcześniejszych) wersji tego artykułu oraz cenne uwagi. Marcinowi Miłkowskiemu należą się zresztą specjalne podziękowania za nieustające wsparcie i zagrzewanie do pracy.

[2] Podobnie jak wielu innych autorów, używam terminu „fizykalizm typów" w odniesieniu do stanowiska w filozofii umysłu, choć np. Fodor [2008] stosuje go w sposób bardziej ogólny, obejmujący także poglądy w filozofii nauk społecznych. „Fizykalizm typów" nie jest jedyną nazwą omawianej koncepcji w filozofii umysłu – znamy tę doktrynę również jako „teorię identyczności rodzajowej", „materializm stanu centralnego", „teorię stanu mózgu", a nawet „fizykalizm typiczny"; por. np. Kim [2002] s. 9.

Moja ocena omawianych koncepcji znacznie odbiega od tego, co myślał o nich sam Putnam. Po pierwsze, sądzę, że słaby funkcjonalizm da się utrzymać, a w każdym razie warto się nad nim poważnie zastanowić. Po drugie, argument z wielorakiej realizacji, który miał obalić teorię identyczności, jest wadliwy. Po trzecie, funkcjonalizm jest w najlepszym razie tylko częściowo niezgodny z fizykalizmem typów.

2. SYTUACJA WYJŚCIOWA

W latach sześćdziesiątych ubiegłego wieku materialistycznie nastawieni filozofowie analityczni dzielili się, z grubsza biorąc, na behawiorystów logicznych oraz na zwolenników teorii identyczności.

Behawioryści twierdzili, że *między zdaniami o przeżyciach (stanach umysłu) a zdaniami o zachowaniach zachodzą konieczne związki znaczeniowe*, które sprawiają, że niektóre wypowiedzi, takie jak „Postępowanie Kowalskiego świadczyło o tym, że jest wściekły, ale możliwe, że wcale się nie gniewał", nie mają sensu, nawet jeśli na pierwszy rzut oka brzmią rozsądnie i składnie. Putnam wyróżnia dwie odmiany behawioryzmu logicznego: mocną i słabą; wersja mocna jest wzbogacona o tezę, że *zdania o przeżyciach są przekładalne na zdania o zachowaniach*. Innymi słowy, behawioryzm mocny był stanowiskiem redukcjonistycznym, natomiast słaby – antyredukcjonistycznym.

Teoria identyczności (fizykalizm typów) głosiła, że *terminy mentalne odnoszą się do pewnych, najczęściej nieodkrytych, rodzajów zdarzeń, stanów lub własności ośrodkowego układu nerwowego*[3]. W postaci niesemantycznej brzmiałaby: rodzaje przeżyć są tożsame z typami stanów mózgu. Zgodnie z dominującą tradycją interpretacyjną skupię się na wzmocnionym fizykalizmie typów, choć spreparowanie słabej (nieredukcyjnej) wersji tego poglądu nie wymaga szczególnej pomysłowości – wystarczy się powołać na ubóstwo i brak precyzji języka naturalnego, jak to czynili behawioryści logiczni odrzucający tezę o przekładalności, albo na niewspółmierność teorii. Dlatego do poprzedniego sformułowania dodaję tezę, że *właściwa teoria umysłu jest redukowalna do właściwej teorii mózgu*. Koniunkcję teorii identyczności i tezy o redukowalności nazywam *redukcjonizmem psychofizycznym* lub krótko *redukcjonizmem*.

[3] Chodzi o stany, zdarzenia lub własności *fizyczne*, czyli takie, których obecność można stwierdzić w ekstraspekcji, choć Putnam tego przymiotnika nie dodaje. Zapewne można głosić istnienie niefizycznych stanów ośrodkowego układu nerwowego lub niefizycznych zdarzeń w mózgu – doktryny tego typu nie będą jednak odmianami fizykalizmu typów.

Na użytek tego artykułu przyjmuję, że redukowalność[4] jest własnością teorii rozumianych jako zbiory zdań, choć ujęcie w kategoriach podejścia niezdaniowego samo w sobie nie powinno raczej mieć wpływu na zrozumienie mojego wywodu. Formalnie biorąc, teoria wtórna T_1 jest redukowalna do teorii pierwotnej T_2, jeżeli z koniunkcji T_2 oraz pewnych warunków dodatkowych i postulatów można wyprowadzić wszystkie prawa T_1. Warunki dodatkowe bywają niezbędne, bo często T_1 jest, z punktu widzenia T_2, prawdziwa tylko w przybliżeniu lub tylko w określonych warunkach; wówczas nie sposób wyprowadzić samej T_1 z T_2, można jedynie wyprowadzić obraz T_1. Postulaty natomiast wiążą charakterystyczne (tj. nieobecne w T_2) terminy deskryptywne T_1 ze słownikiem T_2, umożliwiając dedukcję praw teorii wtórnej z teorii pierwotnej; jeśli T_1 nie jest z punktu widzenia T_2 w pełni prawdziwa, wówczas postulaty mają status eksplikacji[5].

Redukcja jest rodzajem wyjaśnienia nomologiczno-dedukcyjnego, dlatego nakłada się na nią rozmaite warunki o charakterze pozaformalnym, które mają na celu zagwarantowanie mocy eksplanacyjnej. Należy o tym pamiętać choćby dlatego, że funkcjonuje w literaturze[6] ujęcie, wedle którego koniecznym i wystarczającym warunkiem redukcji T_1 do T_2 jest istnienie wspomnianych postulatów, nazywanych najczęściej prawami pomostowymi: mają one mieć postać równoważności lub identyczności; dyskutuje się, czy są koniecznie czy przygodnie prawdziwe, czy mają status sądów apriorycznych czy aposteriorycznych. Już Kemeny i Oppenheim słusznie uznają taką redukcję za formalne ćwiczenie pozbawione znaczenia poznawczego[7]. W istocie, najkrótsza rozsądna odpowiedź na pewną wersję argumentu z wielorakiej realizacji „pokazującą", że redukcja psychologii do nauk o mózgu jest niemożliwa, bo nie będzie odpowiednich praw pomostowych, polegałaby na zwróceniu uwagi na ten prosty fakt: prawa pomostowe nie

[4] Wyraz „redukcja" jest wieloznaczny, więc ilekroć chcemy powiedzieć coś ogólnego na temat związków między teoriami, zawsze znajdzie się ktoś, kto będzie się domagał jego definicji, najlepiej osadzonej w odpowiednio bogatej typologii. Nie twierdzę, że postulat ten jest niesłuszny; twierdzę natomiast, że trudno mu sprostać. Stosowanie etykiet, takich jak „redukcja derywacyjna", niewiele samo w sobie tłumaczy, bo i one są wieloznaczne - „derywacja" to albo tyle co „dedukcja" (zob. np. Poczobut [2009] s. 136), albo tyle co „różne matematyczne techniki wyprowadzania jednych formuł z drugich, np. na drodze przechodzenia do granicy lub pomijania dalszych wyrazów szeregu rozwinięcia pewnych funkcji" (Strawiński [1997] s. 117). Sprawę dodatkowo komplikują pytania o adekwatność teorii redukcji, o znaczenie redukowalności teorii dla ontologii i epistemologii, a także związane z nimi zagadnienia dotyczące pojęcia wyjaśniania czy struktury teorii naukowych.

[5] Klasyczne ujęcia redukcji podają Nagel [1970]; Kemeny, Oppenheim [1956]. Przegląd rozmaitych koncepcji w tej dziedzinie można znaleźć np. w Strawiński [1997]; Paprzycka [2005].

[6] Zob. szczególnie Fodor [2008].

[7] Zob. Kemeny, Oppenheim [1956] s. 11.

muszą być nawet równoważnościami (poza tym ich formułowanie jest ostatnim, najmniej istotnym etapem redukcji – ważnym tylko z punktu widzenia logików[8]).

Zanim na scenę wkroczył funkcjonalizm, sytuacja dialektyczna wyglądała, w dużym uproszczeniu, następująco: z jednej strony behawioryzm logiczny był kontrintuicyjny, bo sugerował, że subiektywne stany umysłu nie istnieją (mówiąc ściślej, nawet gdyby istniały, byłyby naukowo nierelewantne – behawioryzm logiczny, jako doktryna semantyczna, nie mówi o świecie wprost, lecz właśnie sugeruje), z drugiej strony stała za nim solidna metodologia oraz przekonująca koncepcja nabywania języka (dziecko nie ma dostępu do przeżyć innych ludzi, więc może się nauczyć poprawnego użycia np. słowa „ból" tylko na podstawie obserwacji zachowań; dlatego subiektywne wrażenie nie może być częścią znaczenia żadnego terminu mentalnego). Redukcjonizm psychofizyczny nie kwestionował natomiast istnienia przeżyć, ujmując je jako stany mózgu, lecz powątpiewano w jego moc eksplanacyjną, a w dodatku trudno go było uzgodnić z panującymi wówczas koncepcjami języka. Ponieważ wszystkie zagadnienia ontologiczne i epistemologiczne rozważano wtedy na płaszczyźnie semantycznej, problem był poważny. Behawioryzm logiczny, dodajmy, sam nie był wolny od tego rodzaju kłopotów. Nie radził sobie zwłaszcza z rozwojem nauki i jego wpływem na język naturalny: gdyby naukowcy potwierdzili np. hipotezę, że ból nie jest tożsamy z określoną konfiguracją bodźców i reakcji, lecz z pobudzeniem pewnego ośrodka mózgu, które może występować niezależnie od grymasu na twarzy, płaczu itd., wówczas behawiorysta, ceniący przecież empirię, byłby zmuszony twierdzić, że hipoteza, o której mowa, nie dotyczy bólu, lecz czegoś innego; zmiana znaczenia wyrazu „ból" w nauce przeniknęłaby jednak szybko do języka potocznego, a behawiorysta straciłby grunt pod nogami – jego teza mogłaby już być prawdziwa jedynie o przeszłości.

3. Słaby funkcjonalizm Putnama

W dwóch wczesnych tekstach poświęconych obliczeniowemu podejściu do umysłu[9] Putnam wydaje się bronić następujących tez, których koniunkcję nazywam *słabym funkcjonalizmem*:

[8] Chodzi o to, że z teorii pierwotnej nie można wydedukować praw zawierających terminy deskryptywne, które w niej nie występują (chyba że jest np. sprzeczna). Z punktu widzenia naukowca trudność dokonania redukcji polega raczej na odkryciu strukturalnych podobieństw obu teorii, czyli na podaniu tego, co nazwałem „warunkami dodatkowymi". Marras [2005] porównuje konieczność sformułowania reguł pomostowych do obowiązku posprzątania po pracy – naukowcy pracują, logicy tylko sprzątają. Warto przy okazji dodać, że istnieją też koncepcje redukcji, które w ogóle obywają się bez reguł pomostowych: zob. Kemeny, Oppenheim [1956] czy Bickle [2006].

[9] Putnam [1960, 1964].

(T1) Skoro można pokazać, że wszystkie klasyczne problemy filozofii umysłu miałyby swoje logiczne odpowiedniki dla robotów (fizycznych realizacji maszyn Turinga), to problemy te są niezależne od jakichkolwiek subiektywnych własności ludzkiego doświadczenia *ani nie mają wiele wspólnego z materiałem, z którego ludzie są zbudowani, tj. zarówno dualizm, jak i teoria identyczności są nierelewantne*[10].

(T2) Kwestie sporne w filozofii umysłu nie mają charakteru faktualnego, lecz konwencjonalny (a może nawet arbitralny) – ich rozwiązanie nie polega na odkryciu, jak się rzeczy mają, lecz na postanowieniu, by traktować je w taki lub inny sposób[11].

3.1. Uzasadnienie (T1)

Żeby uzasadnić (T1), wyobraźmy sobie, za Putnamem, społeczność mobilnych robotów (fizycznych realizacji maszyn Turinga) wyposażonych w elektroniczne receptory odbierające informacje z otoczenia. Maszyny te mają dostęp do niektórych stanów wewnętrznych, co pozwala na formułowanie „sprawozdań pierwszoosobowych" (czegoś w rodzaju komunikatów o błędach, zawartości pamięci itp.).

Ponadto znajdują się mniej więcej na tym samym etapie rozwoju cywilizacyjnego co ludzie pod koniec lat pięćdziesiątych – nie dysponują pełną wiedzą o strukturze świata i swojej architektury; uważają, że teorie naukowe są formalnymi rachunkami częściowo zinterpretowanymi przez reguły korespondencji, że posługiwanie się językiem wymaga milczącej znajomości publicznych kryteriów

[10] Cytuję dwa zdania z początku (Putnam [1960]): „Specifically, I shall try to show that all the issues [that make up the traditional mind-body problem - W.H.; zob. następny przyp.] arise in connection with any computing system capable of answering questions about its own structure, and have thus nothing to do with the unique nature (if it *is* unique) of human subjective experience. [...] Another question connected with the 'mind-body problem' is the question whether or not it is ever permissible to identify mental states and physical states" (zob. Putnam [1975a] s. 362). Łatwo zauważyć, że w świetle podanych cytatów składnik (T1), który zaznaczyłem kursywą, pozostaje słabo udokumentowany. Jest on wyrazem następującej interpretacji (Putnam [1960, 1964]): jeśli klasyczne problemy filozofii umysłu są niezależne od wyjątkowej natury naszego subiektywnego doświadczenia, ponieważ o tych samych zagadnieniach mogłyby dyskutować roboty, których ewentualne subiektywne doświadczenie miałoby zapewne inną naturę, to problemy te są również niezależne od wyjątkowej natury naszych mózgów, ponieważ „mózgi" robotów miałyby zapewne inną naturę. Krótko: celem wspomnianej transformacji zagadnień filozoficznych jest pokazanie, że istotne są tylko aspekty, które pozostają niezmienne ze względu na tę transformację.

[11] Pierwsze zdanie (Putnam [1960]) brzmi: „The various issues and puzzles that make up the traditional mind-body problem are wholly linguistic and logical in character: whatever few empirical 'facts' there may be in this area support one view as much as another" (zob. Putnam [1975a] s. 262).

użycia wyrażeń, że kryteria te mają często charakter empirycznych procedur sprawdzania itd.

Przypuśćmy, że korpus zdań zasadnie uznawanych przez roboty zawiera pewną liczbę twierdzeń korelacyjnych o postaci „Ilekroć jestem w wewnętrznym stanie *A*, tylekroć obwód *X* jest włączony i odwrotnie" (jest to formuła równoważnościowa: maszyny eksperymentalnie stwierdziły współwystępowanie odpowiednich stanów wewnętrznych i działania obwodów, nie zanotowały zaś wyjątków od tych uogólnień). Pierwsza część tego rodzaju wypowiedzi to zdanie obserwacyjne, druga – teoretyczne.

Może się teraz wśród maszyn filozofów wywiązać spór: czy takich korelacji nie da się wyjaśnić, wskazując, że obie części twierdzenia odnoszą się do jednego i tego samego stanu rzeczy – bycie w stanie *A* to nic innego jak posiadanie włączonego obwodu *X*.

Niektóre roboty mogą się takiej identyfikacji sprzeciwiać, argumentując na przykład, że tożsamość jest akceptowalna jedynie wówczas, gdy ma charakter analityczny. Tymczasem to są zdania syntetyczne. (Ten zarzut może paść zarówno z ust dualistów, jak i behawiorystów logicznych).

Niektórzy automatyczni filozofowie będą też utrzymywać, że choć można opisywać jedno zdarzenie lub jeden przedmiot za pomocą kilku nierównoważnych zdań, z *własnościami* jest inaczej. Jeśli wierzyć teoriom semantycznym, które utożsamiają znaczenia z własnościami, różnica znaczeniowa pociąga różnicę ontyczną, a wyrażenie „włączony obwód *X*" i „stan wewnętrzny *A*" są niesynonimiczne (mają odmienne użycie, inne metody sprawdzania). Ten zarzut może sformułować antyredukcjonista wszelkiej maści, choć behawiorysta logiczny musiałby dodatkowo wykazać związki znaczeniowe między słownikiem mentalnym a słownikiem zachowań, na które się powołuje.

Nie będę tych dwóch zarzutów analizował, choć jeszcze do nich wrócę. Przyjrzę się natomiast tzw. *argumentowi lingwistycznemu* (tę nazwę zaczerpnąłem od Putnama).

Zdanie typu „Być w wewnętrznym stanie *A* to tyle co posiadać włączony obwód *X*" nie może w normalnym kontekście wyrażać stwierdzenia – zinterpretowane jako stwierdzenie, byłoby niezrozumiałe. Może natomiast wyrażać następującą konwencję: od dzisiaj zwrot „być w wewnętrznym stanie *A*" wolno za-stępować zwrotem „posiadać włączony obwód *X*" – wtedy jednak zmienia się po prostu znaczenie słów: wypowiedziom o włączonym obwodzie nadaje się interpretację, którą dotychczas przypisywano zdaniom o stanach wewnętrznych. Zatem propozycja robota redukcjonisty jest albo niepoprawna gramatycznie

(i dlatego nie do przyjęcia), albo sprowadza się do zalecenia osobliwych zachowań językowych, pozbawionych wartości eksplanacyjnej.

Na argument lingwistyczny można odpowiadać, przyjmując za dobrą monetę koncepcję języka, na której gruncie został sformułowany. Można mianowicie wskazać, że fakt, iż dana wypowiedź nie ma obecnie standardowego użycia, nie świadczy jeszcze o tym, że nie uzyska go w przyszłości. Język się zmienia, a prorokowanie, że jakaś modyfikacja nie nastąpi, wymagałoby prekognicji. Teza robota redukcjonisty staje się w ten sposób przewidywaniem. Nie jest to stanowisko wygodne, nie sposób go bowiem uzasadnić. Trzeba czekać. Roboty, które mają dłonie podobne do naszych, mogą tylko trzymać kciuki, jeśli są przesądne.

Lepiej pokazać, że rozważane wypowiedzenie ma jednak sens twierdzący. Robot redukcjonista powie, że wyraża ono identyfikację interteoretyczną, która bywa dozwolona i uzasadniona (sensowna), kiedy np. z teorii redukującej da się wywieść dobre przybliżenie praw teorii redukowanej, teoria redukująca pozwala wyjaśnić mankamenty teorii redukowanej lub redukcja pozwala na dokonywanie nowych przewidywań czy na sformułowanie płodnego programu badawczego.

Minimalny zysk z redukcji to oczywiście wykluczenie pewnych pytań jako pozbawionych sensu i uproszczenie ontologii (tj. jeśli się jest realistą – aby uniknąć zbędnych komplikacji, przyjmuję, że wszyscy w tej dyskusji mają nastawienie realistyczne; zakładam w szczególności, że behawioryści logiczni uważają, że zdania o stanach umysłu mogą być prawdziwe, jeśli nie łamią zasad głębokiej gramatyki pojęciowej).

Niestety, druga odpowiedź robota redukcjonisty przypomina pierwszą, co jego adwersarz natychmiast wykorzysta, mówiąc: „Żeby dokonać identyfikacji, o jakiej tu mowa, trzeba mieć dwie teorie (tę redukowaną i tę redukującą), a Ty, redukcjonisto, nie masz bodaj żadnej, a już na pewno nie masz *dostatecznie dobrej* teorii redukującej". Innymi słowy, redukcjonista pozostaje futurystą! Zabobonne roboty anglosaskie krzyżują palec wskazujący z serdecznym...

Zostawmy na chwilę tę dyskusję i spójrzmy na nią z boku. Widzieliśmy, że podstawowym źródłem kłopotów robota redukcjonisty była semantyka oraz fakt, że jego teza dotyczyła przyszłości.

Z drugiej strony – z ludzkiego punktu widzenia – wydaje się, że teoria identyczności stanów wewnętrznych i obwodów jest chybiona, gdyż nie uwzględnia istnienia myślących i czujących systemów, które w ogóle nie są wyposażone w obwody (dlatego redukcjonizm psychofizyczny jest *prima facie* nie do utrzymania).

Co więcej, ludzie dualiści też mają kłopot, bo nawet jeśli ich zarzuty pod adresem redukcjonisty są zasadne, a dadzą się przełożyć na język robotów, to nie-

które maszyny też mogą mieć duszę! W tym sensie tradycyjne odpowiedzi w kwestii umysł-ciało są nierelewantne – por. (T1).

Chęć uniknięcia tej nieprzyjemnej konsekwencji skłaniała dualistów do kwestionowania Putnamowskiej analogii.

Oto dwie standardowe obiekcje, które zgłaszano[12]:

> (O1) Niektóre *qualia* są wewnętrznie przyjemne lub bolesne. Maszyna nie może mieć takich *qualiów*, bo można ją przeprogramować tak, by to, co dotychczas było „przyjemne" (należało do stanów preferowanych), „sprawiało ból" (należało do stanów, których system unika).

Obiekcja ta opiera się, zdaniem Putnama, na wątpliwym założeniu, że istnieją wewnętrznie przyjemne lub przykre *qualia*[13]. Jak to rozumieć? Moja krótka eksplikacja jest następująca: jak wie każdy palacz, *qualia* związane z wdychaniem dymu tytoniowego są najpierw przykre, a potem przyjemne, choć – tu się powołam na własne doświadczenie - sam smak papierosa (*quale*) raczej się nie zmienia. Palenie kolejnych papierosów mimo doznawanej przykrości jest elementem procedury przeprogramowywania, dzięki której niepalący staje się palaczem. To samo można powiedzieć o doznaniach podczas picia alkoholu, jedzenia śledzi w oleju, surowej ryby tudzież dowolnego stałego pokarmu itd. Nie wykazano natomiast istnienia *qualiów*, dla których tego typu modyfikacja byłaby niemożliwa.

Na to ktoś mógłby odrzec, iż *qualia* zmieniają się w zależności od tego, w jakim stanie znajduje się układ nerwowy i inne podsystemy człowieka, a więc mamy tu do czynienia ze zmianą *qualiów*, nie zaś z ich przeprogramowaniem[14]. Jeśli ta obiekcja jest słuszna, to imputowana zmiana *qualiów* jest introspekcyjnie niezauważalna, co samo w sobie osłabia (O1)[15]. Ponadto idea przeprogramowania jakiegokolwiek systemu (czy systemem tym jest człowiek, czy robot) w taki sposób, że stan systemu w ogóle się nie zmienia, jest w najlepszym razie osobliwa,

---

[12] Zob. *ibidem*, s. 390.

[13] *Ibidem*, s. 394: „The other dubious premiss is the existence of *intrinsically* pleasant and painful qualia. This is supposed to be introspectively evident, but I do not find it so". Dziękuję przy okazji recenzentowi ICF „Diametros", który zarzucając, że niewłaściwie rozumiem, co Putnam miał na myśli, zmusił mnie do ponownej lektury omawianych artykułów – dzięki temu znalazłem zacytowane zdanie i uzmysłowiłem sobie, że we wcześniejszej wersji niniejszego tekstu przywłaszczyłem sobie (bezwiednie) pierwszą odpowiedź na (O1).

[14] Tę obiekcję zgłosił anonimowy recenzent ICF „Diametros".

[15] Subiektywne własności przeżyć, które nie są przejrzyste dla introspekcji, wydają mi się bytami nieco paradoksalnymi, tym bardziej że, jak rozumiem, mają się również całkowicie wymykać opisowi w kategoriach trzecioosobowych. Wynika stąd, że wszelkie zdania o takich *qualiach* są z konieczności bardzo słabo uzasadnione.

choćby dlatego, że stan wszystkich układów fizycznych zmienia się i bez programowania[16].

Putnam ma i drugą uwagę pod adresem (O1). Wskazuje mianowicie, że konstrukcja robota może być tak złożona, że jego przeprogramowanie będzie niewykonalne (technicznie lub wręcz fizycznie niemożliwe), a zatem, nawet gdyby pierwsza riposta okazała się nieefektywna, (O1) pozostanie nierozstrzygająca (możliwe, że nie da się przeprogramować ani *qualiów* człowieka, ani *qualiów* robota - analogia nadal działa). Z samym pojęciem *quale* wiąże się zresztą wiele trudności, na których omówienie nie ma tu miejsca[17].

> (O2) Człowiek wypowiada słowa „Widzę czerwień", bo wie, że ma odpowiednie doznanie. Robot nie wie; po prostu generuje pewne dźwięki na skutek zajścia odpowiednich okoliczności.

Ta obiekcja jest bardziej pouczająca – by ją oddalić, słaby funkcjonalista powoła się, po pierwsze, na *zasadę izomorfizmu psychologicznego*: dla dowolnej teorii psychologicznej lub epistemologicznej może istnieć niestandardowy model (np. robot), bo własności psychiczne mają naturę funkcjonalną (tj. są zdeterminowane przez wyróżnione relacje, których członami, w rozmaitych konfiguracjach, są bodziec, stan umysłu i reakcja). Należy przy okazji podkreślić, że akceptacja zasady izomorfizmu psychologicznego jest równoważna uznaniu funkcjonalizmu w standardowym (acz dość szerokim) znaczeniu tego słowa.

Po drugie, warto przyjąć semantykę ról pojęciowych. Dostaniemy wówczas wniosek, że jeśli wszystkie prawa właściwej psychologii są prawdziwe o robotach, to roboty mają umysł.

Strategia Putnama wobec (O2) polega na domaganiu się eksplikacji zdania „Iksiński wie, że ma doznanie *A*". Jeśli krytyk nie dysponuje odpowiednią teorią wiedzy, to jego zarzut będzie pozbawiony treści; jeżeli zaś umie spełnić prośbę Putnama, wówczas ten natychmiast wyczaruje maszynę, o której teoria ta będzie prawdziwa. W rezultacie wydaje się, że (O2) zostaje oddalona bez względu na to, co się stanie.

Żeby dopiec Putnamowi, lepiej byłoby uderzyć bezpośrednio w izomorfizm psychologiczny lub w semantykę ról pojęciowych. Do izomorfizmu jeszcze

---

[16] W świetle interpretacji recenzenta w pełni wysłowiona (O1) musiałaby chyba brzmieć następująco: niektóre *qualia* są wewnętrznie przyjemne lub przykre, choć nie da się tego stwierdzić ani w introspekcji, ani tym bardziej w ekstraspekcji; maszyna nie może mieć takich *qualiów*, bo da się ją przeprogramować, przy czym pod pojęcie przeprogramowania nie podpada żadna procedura, która zmienia stan fizyczny systemu. Odpowiedź: „przeprogramowanie" występujące w (O1) jest nazwą pustą, a istnienia takich *qualiów* nie warto postulować. Tak odczytana, (O1) sama się znosi.

[17] Por. zwłaszcza Shoemaker [1975]; Dennett [1988]; ew. Hensel [2003].

wrócę; co do semantyki, pomysł polegałby na podaniu takiej teorii znaczenia, która nie pozwoli na stosowanie terminów mentalnych do innych systemów niż ludzie oraz pewne zwierzęta, bez względu na ewentualny izomorfizm psychologiczny. Można by do tego celu wykorzystać na przykład odmianę kauzalnej teorii odniesienia nazw naturalnorodzajowych, którą sam Putnam później skonstruował. Na przykład: odniesienie terminu „ból" zostało dawno ustalone podczas chrztu składającego się z wielu zdarzeń wskazywania rozmaitych przedmiotów jako egzemplifikacji bólu oraz nie-bólu, w taki sposób, że żaden człon odpowiedniej relacji tożsamości rodzajowej nie jest własnością przedmiotu nieożywionego; na wszelki wypadek można też wykluczyć zmiany odniesienia lub tak ustalić ich warunki, by nie wystąpiły w interesujących nas przypadkach.

3.2. Uzasadnienie (T2)

Gdy przychodzi do uzasadnienia (T2), rozumowanie Putnama przybiera niespodziewany obrót. Proponuje on mianowicie rozważyć pewną odmianę tzw. problemu innych umysłów, czyli pytanie: czy robotom (psychologicznie izomorficznym do nas) przysługuje świadomość? Przy czym – i to jest właśnie zaskakujące – założymy, że słowo „świadomość" nie figuruje we właściwej teorii psychologicznej, choć mimo to posiada niepuste odniesienie. Bez tego założenia odpowiedź byłaby banalna.

Tu zaczyna się przegląd argumentów. Nie wdając się w szczegóły, powiem tak: wszystkie racje za stanowiskiem „konserwatywnym" (roboty są nieświadome) są bardzo słabe; wszystkie racje na rzecz podejścia „liberalnego" odwołują się natomiast do izomorfizmu psychologicznego, co zważywszy wyjściowy warunek, jest równie nieskuteczne.

Tak oto Putnam dostaje (T2): problem umysł-ciało oraz problem innych umysłów to pseudoproblemy!

(T2) jest wszakże niedostatecznie uzasadniona. Jeżeli jesteśmy realistami w kwestii umysłu i akceptujemy tezę izomorfizmu psychologicznego, to sytuowanie świadomości poza dziedziną psychologii i przyjmowanie wyjściowego warunku Putnama nie ma raczej sensu. Miałoby z pewnością sens, gdyby świadomość w ogóle nie istniała – wtedy wspomniane zagadnienia byłyby faktycznie pseudoproblemami, ale Putnamowi nie o to chodzi, bo mówi wyraźnie, że trzeba w stosunku do nich podjąć decyzję natury logiczno-językowej. Nieistnienie świadomości byłoby przecież kwestią faktów, a nie konwencji.

Wydaje się, że (T2) ma raczej źródło w przekonaniu, że teoria znaczenia jest konwencjonalna, albo nawet mocniej – arbitralna. Teza o *konwencjonalności* semantyki głosiłaby, że w ostatecznym rozrachunku wybór teorii semantycznej ma charakter decyzji – z tego jednak nie wynika, że decyzja ta nie podlega racjonalnej

krytyce: dopuszcza się, a nawet zakłada, że względy pragmatyczne istotnie zawężają pole manewru. Być może faworyzują tylko jedną teorię. Z przeświadczenia o *arbitralności* będzie natomiast wynikała niemal nieograniczona swoboda wyboru zasad teorii semantycznej, czyli sceptycyzm.

Kiedy jeszcze raz spojrzymy na opisane tu obiekcje robota antyredukcjonisty, zauważymy, że wszystkie argumenty, począwszy od argumentu z analityczności zdań o tożsamości, przez odwołanie się do teorii identyfikujących własności ze znaczeniami, po argument lingwistyczny, odsyłają nas do jakiejś koncepcji języka. Zatem opowiedzenie się za którymkolwiek ze stanowisk w filozofii umysłu będzie zależało od rozstrzygnięć semantycznych. Konwencjonalność teorii znaczenia pozwalałaby na przeniesienie rozważań na grunt filozofii języka, arbitralność teorii języka obnażałaby jedynie pozorność podejmowanych problemów. W grę wchodzą, moim zdaniem, obie możliwości, choć osobiście skłaniam się ku drugiej – uzasadnienie tej skłonności wykracza jednak poza ramy tego artykułu.

4. MOCNY FUNKCJONALIZM

Kłopot z tezą o arbitralności semantyki polegał na tym, że stała w sprzeczności z ówczesnym programem filozoficznym Putnama[18]. Nic więc dziwnego, że w 1967 r. zmienił front, przechodząc na pozycje *mocnego funkcjonalizmu*. Centralna intuicja tego stanowiska sprowadza się do traktowania serio nie tylko izomorfizmu psychologicznego[19] oraz samej psychologii, ale także teorii znaczenia. Klasyczne problemy filozofii umysłu stają się w ten sposób realnymi kwestiami empirycznymi. Zadanie badaczy polega na analizie funkcjonalnej organizacji podmiotów oraz odkryciu mechanizmów, które decydują o tym, że jedne systemy traktujemy jako osoby, a inne nie.

Do realizacji programu badawczego mocnego funkcjonalizmu potrzeba kilku teorii. Minimalnie: psychologii, która poda warunki posiadania umysłu, oraz teorii niższego poziomu, która pozwoli stwierdzić, czy konkretny przedmiot faktycznie realizuje takie a takie funkcje. Innymi słowy, psychologia pozwala sformu-

[18] Świadczą o tym zarówno Putnam [1957], jak Putnam [1975b].

[19] Abstrahuję tu od bardziej szczegółowego ujęcia tezy funkcjonalizmu w kategoriach maszyn Turinga, zależy mi bowiem na ogólności argumentacji. Z tego samego powodu nie definiuję pojęcia „funkcja". Nie sądzę, by fakt istnienia rozmaitych eksplikacji tego wyrazu miał wpływ na moje rozumowanie, choć różnice między nimi mogą być istotne dla bardziej szczegółowych ujęć. Funkcję eksplikuje się zazwyczaj w jeden z trzech sposobów: albo przyjmuje się wykładnię kauzalną (wówczas funkcje można wyróżnić w dowolnym układzie fizycznym), albo biologiczno-historyczną (wówczas funkcję mogą mieć tylko cechy lub struktury organizmów), albo biologiczno-inżynieryjną (funkcje mogą mieć cechy lub struktury organizmów lub maszyn). Dobry przegląd głównych strategii eksplikacyjnych w tej dziedzinie daje Krohs [2009].

łować hipotezę, że Iksiński myśli, a szeroko rozumiana fizyka pozwala tę hipotezę potwierdzić lub obalić, wskazując, że dla organizmów jakaś własność *A* jest tożsama z pewną psychiczną własnością *B*. Mamy tu zatem częściową redukcję.

To sprawia, że mocny funkcjonalista musi się zmierzyć ze wszystkimi zarzutami, z którymi zmagał się dotąd teoretyk identyczności. W szczególności: jak argumentować za redukcjonizmem, jeśli brak odpowiednich teorii?

Warto zauważyć, że warunek istnienia dostatecznie dojrzałych teorii jest zbyt restrykcyjny, bo wyklucza wszelkie redukcjonistyczne programy badawcze. Gdyby obowiązywał, wszystkie udane redukcje byłyby zaledwie produktami ubocznymi rozwoju nauki. Hipotezy redukcyjne bywają chyba sensowne i nie ma powodu, by je z punktu odrzucać.

Wydaje się, że wystarczy, jeśli redukcjonista wskaże racje, które w tej chwili przemawiają za redukcją – np. liczne korelacje między stanami umysłu i zdarzeniami fizycznymi albo fakt, że stany umysłu modyfikują zachowanie, a więc oddziałują na materię. Będzie się musiał ponadto uporać z rozsądnymi zarzutami, szkicując na przykład możliwy scenariusz rozwoju wiedzy, który pokaże, dlaczego obiekcje wobec redukcjonizmu są chybione. Tego scenariusza nie należy traktować zbyt dosłownie; stanowi on ilustrację, a nie sztywną doktrynę.

Co ważne, zarówno redukcjonista, jak i antyredukcjonista wygłaszają tezę o przyszłości, a co za tym idzie, uzasadnienia po obu stronach będą z konieczności pełne luk i mniej lub bardziej pobożnych życzeń.

Jeżeli to, co napisałem, jest bliskie prawdy, to mocny funkcjonalizm i redukcjonizm są nadal w grze. Wypada przeto na koniec zapytać, czym się od siebie różnią i czy istnieją jakieś racje, by przedkładać jeden nad drugi.

Wzmocniony fizykalizm typów, jak pamiętamy, przewiduje utożsamienie określonych rodzajów stanów centralnego układu nerwowego z pewnymi typami stanów umysłu; np. odczuwanie bólu to tyle co pobudzenie takiego a takiego obszaru mózgu. Z punktu widzenia funkcjonalisty identyfikacja umysłu z mózgiem bezpodstawnie wyklucza istnienie myślących bezmózgowców. Argument ten jest powszechnie znany i nosi miano *argumentu z wielorakiej realizacji*.

Putnam stosuje dwie jego odmiany: *empiryczną* i *pojęciową*. Odmiana empiryczna składa się z dwóch kroków:

> (1) wskazania rozmaitych różnic między mózgami (mogą to być różnice wewnątrzgatunkowe albo międzygatunkowe, albo wręcz wewnątrzosobowe),
>
> (2) wygłoszenia hipotezy empirycznej: wydaje się nieprawdopodobne, by biologia (lub inna nauka uznawana w tym kontekście za fundamentalną) odkryła jedną własność, która może przysługiwać odpowiedniej liczbie tak anatomicznie zróżnicowanych organizmów.

Paradoksalnie zarzut ten trafia również w funkcjonalistę. Załóżmy bowiem, że chcemy uzasadnić twierdzenie, że mój kot i ja odczuwamy w danej chwili ten sam rodzaj bólu. Jeśli nie jesteśmy dualistami, to wiemy, że w obu wypadkach ból jest zlokalizowany w mózgu, zatem funkcjonalista musi wykazać, że obu mózgom przysługuje jedna własność, tj. że oba są adekwatnie opisywalne w kategoriach funkcjonalistycznej psychologii. A jeśli tak, to istnieje chociaż jedna fizyczna własność mózgu występująca u przedstawicieli wielu gatunków zwierząt.

Na takie dictum słyszymy zazwyczaj, że u kota mogą być pobudzone F-neurony, a u człowieka C-neurony (z punktu widzenia biologii są to inne struktury mózgu), które mają tę samą charakterystykę funkcjonalną tylko z punktu widzenia psychologii!

Nie najlepsza to replika. Biologia często dostarcza wyjaśnień teleologicznych, więc mnóstwo cech, które wyodrębnia, to własności relacyjne, nierzadko powiązane z zachowaniem organizmu. Nie ma powodu, by uznać relacyjną własność psychiczną za nieredukowalną do relacyjnej własności biologicznej. Jeżeli biologowi uda się opisać mózg człowieka i kota w jednolity sposób – tak by pewne struktury występowały u obu zwierząt – to tym lepiej. Fakt, że na niższym poziomie opisu struktury te są różne, nie ma znaczenia. To samo dotyczy niemal wszystkich struktur czy cech, jakie znamy, włącznie z paradygmatycznymi przykładami własności redukowalnych (ewentualnym wyjątkiem będą cechy na poziomie fundamentalnym: cząstki elementarne właściwie się od siebie nie różnią).

Można to wyrazić, mówiąc, że teoretyk identyczności pobił mocnego funkcjonalistę tą samą bronią, którą słaby funkcjonalista wykorzystał przeciwko swojemu oponentowi. Nikt mu nie zabroni wykorzystać osiągnięć funkcjonalisty dla własnych celów: jest taki poziom opisu mózgu, który pozwala dostrzec, że mózg kota i mózg człowieka bywają w tym samym stanie lub miewają te same struktury czy moduły – biologowi nie tylko wolno korzystać z pojęć psychologicznych, lecz w badaniach mózgu bywa to wręcz nieodzowne.

Bechtel, Mundale [1999] przekonująco pokazują, że obecna praktyka naukowa tak właśnie wygląda – obszary mózgu wyodrębnia się, pomijając wiele różnic, dzięki czemu jedna struktura występuje u wielu gatunków zwierząt. Pojęcia psychologiczne odgrywają tu rolę heurystyczną, gdyż zakłada się, że określone struktury mózgu odpowiadają za określone funkcje całego organizmu, w tym za funkcje nazywane psychicznymi. Wątpliwości Putnama nie mają zatem podstaw. Zarówno teoretyk identyczności, jak i funkcjonalista mogą odetchnąć z ulgą.

Czas na pojęciową wersję argumentu z wielorakiej realizacji. Przyjmuje ona postać następującego eksperymentu myślowego: nawet gdybyśmy wiedzieli, że Marsjanin nie ma mózgu, to – zakładając, że ma odpowiednio złożoną budowę

i zachowuje się w określony sposób – uznalibyśmy go za podmiot. Dlatego redukcjonizm jest fałszywy.

Żeby oddalić ten zarzut, wystarczy wskazać, że jest on eksperymentem myślowym właśnie. Jako eksperyment myślowy, argument z wielorakiej realizacji przyczynia się raczej do zrozumienia *naszego sposobu myślenia o umyśle*, nie przyczynia się natomiast do zrozumienia, czym umysł jest (tylko drugie pytanie interesuje redukcjonistę typów!). Chyba że wszystko, co empirycznie poznawalne, jest również poznawalne introspekcyjnie – to jednak pogląd sprzeczny z doświadczeniem.

Fizyki Galileusza nie da się prawomocnie odrzucić, wyobrażając sobie sytuację, w której upuszczamy z wieży dwa przedmioty o różnej masie, i twierdząc, że cięższy spadłby prędzej.

Pojęciowy argument z wielorakiej realizacji najlepiej potraktować jako wnioskowanie do najlepszego wyjaśnienia[20]. Funkcjonalizm głoszący, że natura stanów umysłu jest relacyjna, ma być najlepszym wyjaśnieniem naszych intuicji psychologicznych. Zwłaszcza głębokiego przeświadczenia, że stany umysłu mogą występować również u bezmózgowców.

Jak przekonuje Ramsey, istnieją jednak inne koncepcje, które mogą intuicję wielorakiej realizowalności wyjaśnić, na przykład dualizm. Znamy też eksperymenty myślowe i wierzenia, z którymi funkcjonalizm sobie nie radzi. Wymieńmy kilka: eksperyment Blocka z Chińczykami symulującymi działanie mózgu[21], zombie Chalmersa[22], dziecięce przeświadczenie o podmiotowości zabawek, wiara w drewniane figury tudzież duchy przodków – obdarzone nie tylko potężnymi umysłami, ale również mocami sprawczymi. Rzeźbiarz znający strukturę swego dzieła może wierzyć, że figura Marii, którą wykuł w marmurze lub odlał z brązu, myśli i czuje. Nawet jeśli wiemy (lub założymy), że jakiś przedmiot jest odpowiednio wewnętrznie zorganizowany (por. chińskie społeczeństwo symulujące przez pewien czas pracę ośrodkowego układu nerwowego Blocka), nie przypisujemy mu automatycznie przeżyć (narodowi chińskiemu nie przysługują, zdaniem wielu komentatorów, stany umysłu Neda Blocka).

Z funkcjonalizmu (zasady izomorfizmu psychologicznego) nie sposób wyprowadzić pewnych zdań o konkretnych intuicjach, które wielu z nas żywi, więc nie stanowi on ich najlepszego wyjaśnienia (redukcjonizm psychofizyczny wypada w tej konkurencji jeszcze gorzej). W tej dziedzinie największą mocą eksplana-

[20] Por. Ramsey [2006].

[21] Por. Block [1980].

[22] Por. Chalmers [2010].

cyjną może się poszczycić dualizm. To jednak nie ma znaczenia, jeśli zależy nam na wyjaśnieniu umysłu, a nie poglądów na jego temat.

Wracając do sporu między mocnym funkcjonalizmem a redukcjonizmem psychofizycznym, wniosek, jaki się narzuca, jest następujący: ponieważ oba argumenty z wielorakiej realizacji okazały się nieskuteczne, odrzucenie redukcjonizmu na ich podstawie było błędem.

***

Oczywiście, mocny funkcjonalizm nadal wydaje się nie do pogodzenia z fizykalizmem typów, gdy rozważamy status bezmózgowców: według pierwszego stanowiska psychologicznie izomorficzny do ludzi Marsjanin myśli, według drugiego – niekoniecznie. Warto zauważyć, że odkrycie takiego Marsjanina doprowadziłoby najprawdopodobniej do rewolucji w biologii, której dziedzinę należałoby rozszerzyć na nieziemskie formy życia (w takiej sytuacji redukcja psychologii marsjańskiej do biologii nie jest wcale wykluczona). Jeśli jednak Marsjanin nie jest psychologicznie izomorficzny do nas, a mimo to spontanicznie traktujemy go jako podmiot, wówczas oba poglądy mają kłopot (funkcjonalista mógłby twierdzić, że przyszła psychologia obejmie i takiego Marsjanina; jeśli tak będzie, to redukcjonista odpowie, że wróciliśmy do wcześniejszego przypadku - naszego Marsjanina może wszak objąć i biologia).

Pozostaje jeszcze kwestia szczególnych bezmózgowców, czyli robotów: według funkcjonalisty robot psychologicznie izomorficzny do ludzi myśli, według redukcjonisty – raczej nie, trudno bowiem przypuszczać, że inżynieria stanie się częścią biologii. Kto ma rację?

W tym wypadku spór jest, moim zdaniem, czysto werbalny. I to z kilku powodów. Po pierwsze, scenariusz redukcji do biologii, którym się posługiwał fizykalista typów, należy traktować jako ilustrację. Nowe dane empiryczne skłonią redukcjonistę do wyboru bardziej fundamentalnej teorii redukującej, np. fizyki (robot jest układem fizycznym). Po drugie, swą moc eksplanacyjną funkcjonalizm czerpie wyłącznie z redukcji. Co bowiem sprawia, że Iksińskiego boli noga? Sprawia to realizator bólu. Bez odpowiedzi na pytanie, skąd wiadomo, co jest w tym wypadku realizatorem bólu i jakie własności o tym decydują, nie ma mowy o wyjaśnieniu. Funkcjonalizm bez redukcji jest eksplanacyjnym pustosłowiem. Po trzecie wreszcie, nie wiadomo, czy własności fizyczne świata nie wykluczają skonstruowania systemów myślących z materiałów niebiologicznych.

W praktyce mocny funkcjonalizm kolidowałby z fizykalizmem typów, tylko gdyby dopuszczał istnienie przedmiotów niefizycznych o określonej organizacji funkcjonalnej. To jednak byłoby niezgodne z naturalizmem (trudno zresztą

powiedzieć, co by znaczyło twierdzenie, że obiekt niefizyczny realizuje jakieś funkcje). Poza tym redukcjonizm i funkcjonalizm stanowią jeden pogląd. Innymi słowy, funkcjonalizm jest redukowalny do teorii identyczności typów.

Zdaniem niektórych różnica między funkcjonalizmem a fizykalizmem typów jest wyraźna. Powiedzą, że dla mocnego funkcjonalizmu wystarcza brak identyczności funkcji (umysłowej) i realizującej jej struktury: myślenie (funkcja) nie jest identyczne z mózgiem ani żadnym jego składnikiem. Rozmaite funkcje biologiczne czy umysłowe mają wielorakie bazy realizacji, z którymi nie są identyczne - z tego nie wynika ani dualizm, ani fizykalizm.

Takie postawienie sprawy jest równoznaczne z imputowaniem redukcjoniście prostych błędów kategorialnych. Tymczasem nie trzeba funkcjonalizmu, by błędów tych uniknąć. Redukcjonista dostrzega różnicę między oddychaniem a układem oddechowym, oddychanie nadal jednak ujmuje jako proces fizyczny (nie duchowy, nie neutralny metafizycznie); teza redukcjonisty psychofizycznego utożsamia rodzaje przeżyć z typami fizycznych *zdarzeń* w mózgu, nie zaś z samym mózgiem czy jego strukturami. Oczywiście, struktury są tu niezwykle ważne, bo to one umożliwiają myślenie, ale myślenie nie staje się przez to neutralne metafizycznie.

Inny typowy zarzut pod adresem mojego ujęcia jest następujący. Istnieje stanowisko, zwane funkcjonalizmem realizatorów, utożsamiające własności umysłowe z własnościami niższego rzędu: np. „ból" odnosi się do pobudzenia C-neuronów, gdy mowa o człowieku, gdy zaś mowa o Marsjaninie, „ból" denotuje, powiedzmy, wzrost ciśnienia w hydromauronach (zdarzenie to jest funkcjonalnym odpowiednikiem pobudzenia C-neuronów u człowieka). Istnieje jednak i pogląd przeciwny, zwany funkcjonalizmem ról, według którego „ból" denotuje określone miejsce w organizacji funkcjonalnej, i nie należy go utożsamiać z konkretnym realizatorem.

Wydaje się, że uznawanie funkcjonalizmu ról jest podyktowane przeświadczeniem, iż język odzwierciedla strukturę świata, wobec czego nie wolno nigdy utożsamiać własności desygnowanych przez niesynonimiczne lub niekoekstensywne wyrażenia. Bezpieczniej jest posługiwać się słowem „realizować"... Odłóżmy na bok problem wykluczenia przyczynowego, z którym funkcjonalista ról musi sobie jakoś poradzić; zapomnijmy na chwilę o braku uzasadnienia dla tezy o kopiowaniu ostatecznej struktury bytu przez język naturalny; zamknijmy w szufladzie brzytwę Ockhama. Proponuję ograniczyć się do przypomnienia, że konsekwentne przestrzeganie maksymy o nieutożsamianiu własności denotowanych przez nierównoznaczne wyrażenia nie pozwoli raczej na identyfikację stanów umysłu z rolami przyczynowymi: skoro słowo „ból" nie znaczy w języku natural-

nym tyle co „określone miejsce w sieci powiązań przyczynowych", lecz odnosi się do pewnego *quale* (zob. moje odparcie pojęciowego argumentu z wielorakiej realizacji). Każdy funkcjonalizm, który powołuje się na empiryczną psychologię jako właściwą teorię umysłu, musi usunąć tę barierę semantyczną i staje się redukcjonizmem (zob. dyskusja o indentyfikacjach interteoretycznych wśród robotów filozofów)[23].

Zapytajmy na koniec, dlaczego dostrzeżenie braku wyraźnej różnicy między funkcjonalizmem a redukcjonizmem zajęło nam tak dużo czasu (u mnie trwało to jakieś dziesięć lat, wielu moich kolegów nadal uważa, że ów fakt jest pseudofaktem).

Złożyło się na to kilka niezależnych przyczyn. Po pierwsze, brak oddzielenia empirycznego wątku argumentu wielorakiej realizowalności od wątku pojęciowego doprowadził do scalenia dwóch niezależnych sfer. Dzięki temu można było twierdzić, że dualizm nie jest najlepszym wyjaśnieniem czegokolwiek, bo jest fałszywy. Funkcjonalizm był najbardziej intuicyjnym (zdroworozsądkowym) stanowiskiem, które pozostało na placu boju. Po drugie, funkcjonalizm stanowił syntezę behawioryzmu i teorii identyczności (dwóch dobrych pomysłów), która unikała, jak się wydawało, ich najpoważniejszych wad. Aż chciało się weń wierzyć (filozofowie nader często ulegają myśleniu życzeniowemu). Po trzecie, nieredukcyjna wersja funkcjonalizmu była pojęciowo wygodniejsza (redukcjonizm zawsze się wiązał z technicznymi trudnościami – o ileż łatwiej być antyredukcjonistą). Długo by wymieniać inne czynniki. Zwrócę uwagę na jeszcze jeden.

Otóż pod koniec części 3.1 napisałem, że podważenie Putnamowskiej analogii między ludźmi a robotami wymagałoby uderzenia w jej sedno, czyli w zasadę izomorfizmu psychologicznego lub w semantykę ról pojęciowych. Skonstruowanie konkurencyjnej teorii odniesienia, która zablokowałaby (T1), nie jest trudne. Niestety, obstawanie przy teorii znaczenia, która inaczej traktuje terminy psychologiczne, kiedy odnoszą się do maszyn, a inaczej, gdy się je stosuje do przedstawicieli homo sapiens, nawet jeśli pewne maszyny są funkcjonalnie nieodróżnialne od ludzi, wydaje się nieuzasadnione.

Co więcej, zarówno zasada izomorfizmu psychologicznego, jak semantyka ról pojęciowych wywodzą się z tzw. zastanego poglądu na naturę teorii naukowych (*the received view*). Jeśli bowiem teorie są formułami częściowo zinterpretowanymi za pomocą reguł korespondencji, które wiążą terminy teoretyczne ze

[23] Tzw. funkcjonalizm analityczny, który głosi, że znaczenia terminów psychologii potocznej są relacyjne, jest stanowiskiem semantycznym, niezależnym od mocnego funkcjonalizmu, redukcjonizmu czy dualizmu.

słownikiem obserwacyjnym, a słownik obserwacyjny interpretuje się za pomocą procedur sprawdzania empirycznego, przy czym metody weryfikacji w psychologii nie odwołują się do konkretnych własności fizycznych czy biologicznych, to nic dziwnego, że predykaty takie jak „boli" czy „wie" odnoszą się do pewnych systemów nieożywionych, jeśli systemy te spełniają nakładane przez teorię warunki. Ten bliski związek zasady izomorfizmu psychologicznego z semantyką ról pojęciowych był kolejną przyczyną zwycięstwa mocnego funkcjonalizmu.

Zastany pogląd na strukturę teorii naukowych poddano ostrej krytyce[24]. Powinno to było wpłynąć na kształt wszystkich ważniejszych sporów w filozofii umysłu, zwłaszcza wokół redukcjonizmu, lecz badacze w tej dziedzinie okazali się zaskakująco konserwatywni (w Polsce jednym z niewielu wyjątków jest Katarzyna Paprzycka[25]). Jeżeli teorie nie są formalnymi rachunkami częściowo zinterpretowanymi empirycznie, to nawet myśliciele przywiązani do idei jedności nauki mogą zrezygnować z postulatu redukcji, poprzestając na jakiejś słabszej formie integracji interteoretycznej. Wiele naturalnie zależy od ogólnej wizji nauki, a tych jest co najmniej kilka. Dlatego niniejszy tekst jest z jednej strony tylko historyczny, a z drugiej nadal aktualny.

Bibliografia

Bechtel, Mundale [1999] – W. Bechtel, J. Mundale, *Multiple Realizability Revisited: Linking Cognitive and Neural States*, „Philosophy of Science" (66) 1999, s. 175-207.

Bickle [2006] – J. Bickle, *Reducing Mind to Molecular Pathways: Explicating the Reductionism Implicit in Cellular and Molecular Neuroscience*, „Synthese" (151) 2006, s. 411-434.

Block [1980] – N. Block, *Troubles with Functionalism*, [w:] *Readings in Philosophy of Psychology*, t. 1, N. Block (red.), Harvard University Press, Cambridge, MA 1980.

Chalmers [2010] – D. Chalmers, *Świadomy umysł*, tłum. M. Miłkowski, PWN, Warszawa 2010.

Dennett [1988] – D.C. Dennett, *Quining Qualia*, [w:] *Consciousness in Modern Science*, A. Marcel, E. Bisiach (red.), Oxford University Press, Oxford 1988.

Fodor [2008] – J.A. Fodor, *Nauki szczegółowe (albo: niejednorodność nauki jako hipoteza robocza)*, tłum. M. Gokieli, [w:] *Analityczna metafizyka umysłu. Najnowsze kontrowersje*, M. Miłkowski, R. Poczobut (red.), IFiS PAN, Warszawa 2008, s. 56-75.

Hensel [2003] – W.M. Hensel, *W poszukiwaniu* qualiów, „Przegląd Filozoficzno-Literacki" (6) 2003, s. 15-32.

[24] Zob. zwłaszcza Suppe [1974].

[25] Por. Paprzycka [2005].

Kemeny, Oppenheim [1956] – J. G. Kemeny, P. Oppenheim, *On Reduction*, „Philosophical Studies" (7) 1956, s. 6-19.

Kim [2002] – J. Kim, *Umysł w świecie fizycznym*, tłum. R. Poczobut, IFiS PAN, Warszawa 2002.

Krohs [2009] – U. Krohs, *Functions as Based on a Concept of General Design*, „Synthese" (166) 2009, s. 69-89.

Marras [2005] – A. Marras, *Consciousness and Reduction*, „British Journal for the Philosophy of Science" (59) 2005, s. 335-361.

Nagel [1970] – E. Nagel, *Struktura nauki*, tłum. Giedymin, Rassalski, Eilstein, PWN, Warszawa 1970.

Paprzycka [2005] – K. Paprzycka, *O możliwości antyredukcjonizmu*, Semper, Warszawa 2005.

Poczobut [2009] – R. Poczobut, *Między redukcją a emergencją. Spór o miejsce umysłu w świecie fizycznym*, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław 2009.

Putnam [1957] – H. Putnam, *Psychological Concepts, Explication and Ordinary Language*, „The Journal of Philosophy" (54) 1957, s. 94-99.

Putnam [1960] – H. Putnam, *Minds and Machines*, [w:] *Dimensions of Mind*, S. Hook (red.), New York University Press, New York 1960; przedruk [w:] Putnam [1975a] s. 362-385.

Putnam [1963] – H. Putnam, *Brains and Behavior*, [w:] *Analytical Philosophy Second Series*, R. Butler (red.), Basil Blackwell, Oxford 1963; przedruk [w:] Putnam [1975a] s. 325-341.

Putnam [1964] – H. Putnam, *Robots: Machines or Artificially Created Life?*, „The Journal of Philosophy" (61) 1964, s. 668-691, przedruk [w:] Putnam [1975a] s. 386-407.

Putnam [1967] – H. Putnam, *Psychological Predicates*, [w:] *Art, Mind and Religion*, W.H. Capitan, D. D. Merrill (red.), University of Pittsburgh Press, Pittsburgh 1967, przedruk jako: *The Nature of Mental States*, [w:] Putnam [1975a] s. 408-428.

Putnam [1975a] – H. Putnam, *Mind, Language and Reality. Philosophical Papers*, t. 2, Cambridge University Press, Cambridge 1975.

Putnam [1975b] – H. Putnam, *The Meaning of* Meaning, [w:] *Language, Mind and Knowledge*, K. Gunderson (red.), University of Minnesota Press, Minneapolis 1975 (polski przekład: *Znaczenie wyrazu „znaczenie"*, tłum. A. Grobler [w:] H. Putnam, *Wiele twarzy realizmu i inne eseje*, PWN, Warszawa 1998, s. 93-184).

Ramsey [2006] – W. Ramsey, *Multiple Realizability Intuitions and the Functionalist Conception of the Mind*, „Metaphilosophy" (37) 2006, s. 53-73.

Shoemaker [1975] – S. Shoemaker, *Functionalism and Qualia*, „Philosophical Studies" (27) 1975, s. 291-315.

Strawiński [1997] – W. Strawiński, *Jedność nauki, redukcja, emergencja*, Aletheia, Warszawa 1997.

Suppe [1974] – F. Suppe, *The Search for Philosophic Understanding of Scientific Theories*, [w:] *The Structure of Scientific Theories*, F. Suppe (red.), University of Illinois Press, Urbana 1974.