

Hilary Putnam

## **Argument teoriomodelowy a poszukiwanie realizmu zdroworozsądkowego<sup>1</sup>**

Poproszono mnie o scharakteryzowanie „sporu o realizm” i mojego obecnego podejścia do tych kwestii. Z przyjemnością podejmuję się tego zadania, ponieważ wydaje mi się, że chociaż większość moich krytyków i zwolenników zdołała dostrzec, iż są to kwestie skomplikowane i głębokie, to sędzę, że krytycy ciągle nie w pełni docenili, jak bardzo są one skomplikowane i głębokie. W każdym razie w pierwszym paragrafie artykułu przedstawię bardzo skondensowaną historię ewolucji moich poglądów i krótko omówię ten rodzaj realizmu, który obecnie preferuję. W trakcie tej ewolucji, w ciągu ostatnich dwudziestu lat z okładem, doszedłem do porzucenia „anty-realizmu” i „semantyki weryfikacjonistycznej”, których broniłem w takich esejach, jak *Realism and Reason* (1977) i *Models and Reality* (1980) oraz w mojej książce *Reason, Truth and History* (1981). Stało się to wszelako nie z powodu rzekomych błędów w moim „argumentie teoriomodelowym” przeciwko realizmowi metafizycznemu. Jestem teraz bowiem przekonany, że chociaż argument teoriomodelowy nie był błędny, to niektóre założenia na których się opierał — założenia, które były niemal powszechne w filozofii po Kartezjuszu — muszą zostać odrzucone. Jednak założenia te nigdy nie były kwestionowane także przez moich krytyków.

---

<sup>1</sup> Tytuł oryginału: *The Model-Theoretic Argument and the Search for Common Sense Realism*. Jak dotąd tekst ukazał się jedynie w przekładzie niemieckim: *Das modeltheoretische Argument und die Suche nach dem Realismus des Common sense*, [w:] *Realismus*, hrsg. M. Willaschek, Paderborn 2000, Ferdinand Schöningh, s. 125-142. Przekład za zgodą i na podstawie oryginału udostępnionego przez Autora. Oryginał ten został opatrzony przez Prof. Putnama dodatkiem *What I Meant by „Common Sense Realism”*, napisanym w 2009 roku specjalnie na potrzeby polskiego przekładu [przyp. tłum.].

W drugim paragrafie przedmiotem rozważań (i obrony) będzie ta część argumentu, którą moi krytycy niemal powszechnie uznali za błędną, tj. krok „jeszcze więcej teorii”. Natomiast w paragrafie końcowym poczynię kilka uwag ogólnych, dotyczących całej problematyki.

### KRÓTKA HISTORIA MOICH POGLĄDÓW

Gdzieś we wczesnych latach siedemdziesiątych zacząłem zastanawiać się nad tym, co ostatecznie określiłem jako „argument teoriomodelowy”. Opisuję i analizuję ten argument w drugim paragrafie niniejszego eseju, lecz teraz niech mi będzie wolno przedstawić jego końcowy rezultat. Argument jest skierowany przeciwko „realiście metafizycznemu”. Według mojej charakterystyki tego oponenta (którego przedstawiam jako moje poprzednie wcielenie), realista metafizyczny jest przekonany, że prawda zakłada korespondencję z rzeczywistością, ale w zdecydowanie post-tarskiańskim sensie „korespondencji”. Post-tarskiański „realista metafizyczny” przyjmuje relację zwaną *odniesieniem*, zachodzącą pomiędzy każdą nazwą własną<sup>2</sup> w dowolnym języku a bytem, który ona denotuje; pomiędzy każdym jednoargumentowym predykatem języka a każdym elementem ekstensji tego predykatu; pomiędzy każdym dwuargumentowym predykatem a każdą uporządkowaną parą z ekstensji tego predykatu; *etc.*<sup>3</sup> Prawda (jako predykat orzekany o zdaniach tego języka) ma być tu definiowana w kategoriach odniesienia, za pomocą metod Tarskiego. Tak więc, ściśle biorąc, w tego rodzaju „teorii korespondencyjnej” prawda jest *definiowana w kategoriach* korespondencji (pomiędzy predykatami a elementami ich ekstensji), nie jest zaś traktowana jako coś, co *polega na* korespondencji (pomiędzy zdaniami i „rzeczywistością”, albo zdaniami i „światem”, albo zdaniami i „stanami rzeczy”).<sup>4</sup>

Realista metafizyczny jest ponadto przekonany, że przedmioty z ekstensji różnych predykatów są niezależne od umysłu, iż korespondencja jest dokładnie jedna (i w rzeczy samej ma zastosowanie do *wszystkich* języków i teorii jednocześnie) oraz

<sup>2</sup> Rzecz jasna, ponieważ język jest tu ujmowany jako sformalizowany w sposób standardowy, dopuszczalne są tylko nazwy własne, mające denotacje.

<sup>3</sup> A zatem „odniesienie” i relacja, którą Tarski nazywał „spełnianiem”, są wzajemnie definiowalne.

<sup>4</sup> Podkreślam ten post-tarskiański charakter „korespondencyjnej teorii prawdy”, którą rozważam w moich artykułach, ponieważ jest to notorycznie pomijane przez krytyków. Innym powszechnym błędem jest przedstawianie tych artykułów jako *całkowicie* „odrzucających korespondencyjną teorię prawdy”. Pomija się wówczas fakt, iż odrzucam tę teorię tylko wtedy, gdy „odniesienie” ma być relacją *niezależną od teorii*, jakąś jedną metafizyczną relacją korespondencji, zachodzącą pomiędzy dowolnym terminem dowolnego języka a odpowiednim podzbiorem stałej *całości* „wszystkich bytów”. Nigdy jednak nie zaprzeczałem temu, że można mówić o korespondencji pomiędzy terminem opisowym a bytami, do których się on odnosi, kiedy pojęciom „korespondencji” i „odniesienia” nadaje znaczenie sposób, w jaki używamy ich w *konkretnym kontekście*. Realiści metafizyczni na ogół z góry przesądzają kwestię (przez zwykle ignorowanie jej), czy „odniesienie” jest *samo w sobie czymś kontekstualnym, czy „absolutnym”*.

że prawda i korespondencja, na której prawda się opiera, są całkowicie *nie-epistemiczne*, czyli nawet teoria, która jest optymalna pod kątem wszystkich rozsądnych standardów epistemicznych, może faktycznie być *falszywa*. Argument teoriomodelowy, który przedstawię w następnym paragrafie, przekonał mnie na długi czas, że jest to pogląd niespójny. Z grubsza rzecz ujmując, rezultat argumentu jest taki, że jeśli teoria jest epistemicznie idealna, to w pewnym sensie istnieją gwarancje, iż jej predykaty rzeczywiście się odnoszą, a jej zdania są „rzeczywiście prawdziwe” — idea, że epistemicznie idealna teoria może być *falszywa* jest niespójna.

Pod wpływem tej konkluzji zostałem zmuszony do poszukiwania bardziej „epistemicznej” koncepcji prawdy, koncepcji, w której epistemicznie idealna teoria ma zagwarantowaną słusność.

W książce *Reason, Truth and History* (1981) zaproponowałem taką koncepcję. Utrzymywałem tam, że twierdzenie jest prawdziwe wtedy i tylko wtedy, gdy byłoby racjonalnie akceptowalne w idealnych warunkach poznawczych; ponieważ jednak, jak wskazywałem, „idealne warunki poznawcze” nie mogą być faktycznie osiągnięte, tak samo jak w świecie nie można faktycznie znaleźć „powierzchni bez tarcia” (Putnam 1981, s. 55), twierdzenie jest prawdziwe wtedy i tylko wtedy, gdy byłoby racjonalnie akceptowalne w *wystarczająco dobrych* warunkach poznawczych.<sup>5</sup> Uważam, że pokrywa się to dokładnie z tym, co Crispin Wright nazywa „superstwierdzalnością”.<sup>6</sup>

Tę identyfikację prawdy z superstwierdzalnością sam zacząłem uważać za nieco niewiarygodną, lecz w tym czasie (w późnych latach siedemdziesiątych i wczesnych osiemdziesiątych) nie wiedziałem, jak — zważywszy na klęskę realizmu metafizycznego — nadać sens pojęciu prawdy w jakiś inny sposób.

*Realism and Reason* (1977)<sup>7</sup> zawierał inny, całkowicie odmienny argument przeciwko realizmowi metafizycznemu, argument z fenomenu jednakowo poprawnych teorii (od czasu *The Many Faces of Realism* (1987) używałem terminu „względność pojęciowa” dla tego fenomenu), które na pierwszy rzut oka okazują się ze sobą niezgodne, lecz obie są „słuszne”. Czysta matematyka jest pełna przykładów (np. teoria mnogości, w której funkcje są utożsamione ze zbiorami uporządkowanych par oraz teoria mnogości, w której zbiory są utożsamione z „funkcjami charakterystycznymi”, są ze sobą niezgodne w takim sensie, iż ich połączenie jest sprzeczne, lecz *żaden* logik nie przypuszcza, że istnieje „obiektywny fakt” (*fact of the matter*), rozstrzygający o tym, która z nich jest „rzeczywiście słuszna”). Są też na to przykłady w teorii

<sup>5</sup> Zob. ostatni rozdział mojej książki *Representation and Reality* (1988).

<sup>6</sup> Wright jednak tak nie myśli! (Ponieważ mylnie odczytuje on *Reason, Truth and History*). Zob. Wright 1992, s. 44 nn.

<sup>7</sup> Był to mój wykład przewodniczącego (*Presidential Address*) na zjeździe Wschodniego Oddziału Amerykańskiego Towarzystwa Filozoficznego w 1976 roku. Po raz pierwszy przedstawiłem w nim swój argument teoriomodelowy oraz wyrzekłem się realizmu metafizycznego. Został on włączony do mojej książki *Meaning and the Moral Sciences* (1978).

przestrzeni i czasu (teoria, w której punkty są indywiduami oraz teoria, w której punkty są zbieżnymi zbiorami kul), a nawet w fizyce.

Jednakże oba argumenty przeciwko realizmowi metafizycznemu są wymierzone w zupełnie odmienne składniki tego stanowiska. Argument teoriomodelowy był skierowany przeciwko twierdzeniu, że nawet epistemicznie idealna teoria może być fałszywa, a stąd jego ostateczny rezultat był weryfikacjonistyczny. Argument z fenomenu względności pojęciowej był skierowany przeciwko twierdzeniu, iż (poza trywialnym przypadkiem, w którym jakiś termin jest używany w odmiennym znaczeniu w dwóch „niezgodnych” teoriach<sup>8</sup>) sama natura tego, o czym mówimy musi wyznaczać *dokładnie jedną* prawdę na ten temat. Używając bardziej precyzyjnego języka, realista metafizyczny przyjmuje, że świat (ograniczam się tu do twierdzeń empirycznych) musi wyznaczać (1) stałą całość „przedmiotów indywiduowych” oraz (2) stałą całość „własności i relacji” tych przedmiotów indywiduowych (i odpowiednio dla wyższych typów). Oznacza to, że jeśli wyobrazimy sobie język z jedną nazwą własną dla każdego indywiduum i jednym predykatem dla każdej własności i relacji, *etc.* (aż do pewnego ustalonego typu w znanej russellowskiej hierarchii), to w tym języku „infinitystycznym” będzie istniała *dokładnie jedna prawdziwa teoria świata*.

Natomiast akceptacja względności pojęciowej jako rzeczywistego i nietrywialnego (i niedającego się *strywializować*) fenomenu oznacza akceptację, iż *świat nie wyznacza jakiegось jednego uprzywilejowanego użycia nawet takich „terminów logicznych”* jak „przedmiot”, „własność” lub „istnieje”. Całe poszukiwanie odpowiedzi na pytanie o „rzeczywistą ontologię” świata jest z gruntu źle pomyślane. Zauważmy, że akceptacja tego nie wymaga akceptacji weryfikacjonizmu, który jest rezultatem mojego „argumentu teoriomodelowego”. Zgodnie z moim obecnym stanowiskiem *należy* uznać względność pojęciową za rzeczywisty i nietrywialny fenomen, lecz trzeba odrzucić założenia, na których opierał się argument teoriomodelowy.

Ponieważ miałem pewne wątpliwości co do identyfikacji prawdy z superstwierdzalnością, w moich wykładach im. Carusa (Putnam 1987), ograniczyłem swój atak na realizm metafizyczny do obrony względności pojęciowej, zachowując milczenie na temat argumentu teoriomodelowego i utożsamienia prawdy z „wyidealizowaną racjonalną akceptowalnością”. Niemniej ostatni rozdział *Representation and Reality* (1988) nadal tego utożsamienia bronił, podkreślając jednak jego aspekt realistyczny.

Mniej więcej w tym czasie odbyła się w Madrycie konferencja na temat mojej filozofii, na której był obecny Cesar Gomez, jeden z najważniejszych światowych fizyków matematycznych i dobry filozof „amator”. Podczas dyskusji nad moim argumentem teoriomodelowym na tej konferencji Gomez poczynił istotną uwagę. Zasugerował on, że mój argument teoriomodelowy zakłada kartezjański pogląd na per-

<sup>8</sup> Czasami argumentuje się — i nie robią tego tylko realisci metafizyczni — że wszystkie przypadki teorii, które są słuszne, lecz na pierwszy rzut oka niezgodne ze sobą, okazują się w następujący sposób trywialne: pewien termin posiada inne znaczenie (w zwykłym sensie wyrażenia „inne znaczenie”). Odpowiadam na tę tezę w Putnam 1994b.

cepcję, pogląd zgodnie z którym nasze treści percepcyjne (*percepts*) są przedmiotami „wewnątrz” nas, przedmiotami w „teatrze umysłu”; powiedział też, iż „być może *Sense and Sensibilia* Johna Austina zawierają sposób pozwalający na rozwikłanie całego tego problemu”. Nie zgodziłem się wówczas, argumentując, że „nawet gdyby Austin miał rację, nie pomogłoby to w tej kwestii”, lecz po kilku latach doszedłem do przekonania, iż Gomez miał całkowitą rację! (Fakt, iż zajmowałem się filozofią Williama Jamesa, który był zdecydowanym obrońcą „realizmu bezpośredniego”, odegrał w tym znaczącą rolę — zob. Putnam 1996).

Aby dostrzec znaczenie realizmu bezpośredniego dla argumentu teoriomodelowego, musimy zastanowić się nad pojęciem „epistemicznie idealnej” teorii, którym ten argument się posługuje. „Epistemicznie idealna” teoria z założenia powinna nie tylko spełniać „wymogi teoretyczne”, takie jak optymalna prostota, elegancja, koherencja z wcześniejszymi doktrynami, wiarygodność, *etc.*, lecz także w maksymalny sposób spełniać nasze „wymogi operacyjne”. Ale czym jest „wymóg operacyjny”?

Jeśli „wymóg operacyjny” to wymóg o postaci, którą ilustruje następujący przykład: „Zazwyczaj, kiedy teoria T implikuje zdanie «Hilary Putnam widzi krowę» Hilaremu Putnamowi będzie się wydawać, że widzi krowę”, to przyjęcie założenia, iż owe wymogi mogą być spełnione, chociaż wchodzące w grę zdania są „w rzeczywistości fałszywe” w realistycznometafizycznym sensie, byłoby uznaniem, że nasze percepcje są w dużym stopniu błędne — np. jesteśmy „mózgami w naczyniu” lub czymś w tym rodzaju. Wówczas jednak można zapytać osobę przyjmującą takie założenie: „Jak to się dzieje, że uznaje się, iż nasze terminy odnoszą się do takich rzeczy jak krowy, jeśli to nie krowy są tym, co widzimy?”. Do pokazania niespójności tego scenariusza wystarczy przyjąć, że *na odniesienie nałożone są wymogi przyczynowe*.<sup>9</sup> Z kolei argumentowanie (tak jak w „argumencie z permutacji indywiduów” w *Reason, Truth and History*), że chociaż zazwyczaj mamy przed sobą krowy, kiedy wypowiadamy zdanie w rodzaju „Właśnie widzę krowę”, to umysł nadal nie ma sposobu, by oddzielić krowy od wszystkich innych rzeczy, które znajdują się wśród rodowodu przyczynowego „krowich danych zmysłowych” (lub „percepcyjnych reprezentacji”), prowadzących do akceptacji tego zdania, całkowicie opiera się na następującym obrazie: operacje umysłu rozpoczynają się od czegoś takiego jak „*dane zmysłowe*” (wewnętrzne reprezentacje); owe reprezentacje powiązane są z tym, co jest „na zewnątrz umysłu” nie przez jakiegokolwiek procesy poznawcze (wszystkie procesy poznawcze zachodzą „wewnątrz” umysłu), lecz w najlepszym razie przez przyczynowe oddziaływanie sprawcze. Ale ten obraz jest właśnie tym, który realizm bezpośredni — lub lepiej, *realizm zdroworozsądkowy* — atakuje. Dla realisty zdroworozsądkowego nie istnieje problem tego, w jaki sposób umysł ludzki może wyodrębnić przedmiot odniesienia lub klasę przedmiotów odniesienia terminu, przynajmniej jeśli owe przedmioty odniesienia są obserwowalne. Ponieważ w świetle poglądu

<sup>9</sup> Argument przeciwko scenariuszowi mózgow w naczyniu, opierający się na idei, że na odniesienie nałożone są wymogi przyczynowe, jest przedstawiony w rozdziale 1. książki: Putnam 1981.

realisty zdroworozsądkowego umysł percypuje zarazem własne terminy i ich zamierzone przedmioty odniesienia, może on po prostu powiedzieć: „Na czym polega problem?”. Według realisty zdroworozsądkowego relacja pomiędzy naszymi terminami a ich odniesieniami jest sama w sobie *poznawcza*, nie zaś tylko „przyczynowa”.

W rezultacie tych wszystkich refleksji, wygłaszając wykłady im. Deweya w 1994 roku (Putnam 1994a), obrałem tę drogę, do której Cesar Gomez przekonywał mnie dziesięć lat wcześniej. Argumentowałem bowiem za „realizmem zdroworozsądkowym” i przeciwko weryfikacjonizmowi, do którego doprowadził mnie mój argument teoriomodelowy. Argumentowałem także, iż nie ma nadziei na obronę realizmu zdroworozsądkowego, jeśli będziemy obstawać przy „interfejsowej” koncepcji percepcji (i, tak samo, pojęć), którą odziedziczyliśmy po kartezjanizmie. Sądzę, że moi krytycy popełniają ten sam błąd, który ja popełniłem, trzymając się idei, iż możemy mówić o „wymogach operacyjnych” i „wymogach teoretycznych” (tak jak pozytywiści logiczni!) i jednocześnie bronić realizmu.

### OBRONA KROKU „JESZCZE WIĘCEJ TEORII”

Zasadniczym błędnym punktem argumentu teoriomodelowego miał być krok zwany „jeszcze więcej teorii” (*just more theory move*). Jednym z nielicznych komentatorów, którzy *nie* uważają go za błędny, jest Igor Douven (1999). Chociaż, jak się okaże, są momenty, w których jego rekonstrukcja jest nietrafna, to w kluczowych punktach dostrzega on istotę problemu z wyjątkową klarownością. W tym paragrafie zaprezentuję i zbadam jego rekonstrukcję tego argumentu.

Douven rozpoczyna od uwagi (1999, s. 479), że w samym sercu doktryny realizmu metafizycznego, tak jak ją przedstawiam, znajduje się teza nazwana przez niego fallibilizmem metodologicznym (FM), która jest właśnie doktryną, zgodnie z którą „nawet epistemicznie idealna teoria może być daleko od prawdy”. Twierdzi następnie (Douven 1999, s. 479), iż kolejną tezę rozważanego przeze mnie realizmu metafizycznego jest coś nazwanego naturalizmem semantycznym. Sedno tej tezy jest następujące: „semantyka jest nauką empiryczną, taką samą jak każda inna” (Devitt 1996, s. 48). Dalej Douven powiada:

Jedną z konsekwencji tego poglądu jest to, że teorie odniesienia [...] są teoriami empirycznymi i wobec tego muszą być zaakceptowane lub odrzucone na podstawie dokładnie tych samych kryteriów, które decydują o wyborze teorii w obrębie tego, co tradycyjnie uważamy za nauki przyrodnicze (Douven 1999, s. 479).

Będę starał się wyjaśnić (co Douven być może jako pierwszy komentator zauważył), że krytycy argumentu teoriomodelowego nie zdołali spełnić wymogu naturalizmu semantycznego — wymogu, który sami akceptują i którego zalety wychwalają!

W rzeczywistości Douven mógłby w sposób bardziej wyrazisty przedstawić swoją linię argumentacji (którą wkrótce omówię), gdyby nie popełnił pewnego błędu we wcześniejszym artykule, który nie został poprawiony w artykule późniejszym,

obecnie rozważanym, chociaż ten wcześniejszy artykuł jest *prima facie* nie do pogodzenia z konkluzjami późniejszego. W tym wcześniejszym artykule<sup>10</sup> Douven zamierza pokazać, że „byłoby cudem, gdyby istniał świat taki jak nasz, lecz odniesienie byłoby w nim radykalnie niezdeterminowane”.

Argument Douvena we wcześniejszym artykule odwoływał się do postaci argumentu teoriomodelowego, którą przedstawiłem w *Reason, Truth and History*, a w szczególności do sformułowanego przeze mnie argumentu, mającego pokazać, iż fakt, że nasza ewolucja przebiega pomyślnie (do tego stopnia, że „ciągle tu jesteśmy”) *nie* prowadzi do odrzucenia radykalnego niezdeterminowania odniesienia, które w myśl argumentu teoriomodelowego zagraża realności metafizycznemu. Douven prawidłowo referuje, że utrzymuję, iż

jeśli wystarczająco dużo naszych przekonań dyrektywnych [przekonań o postaci „jeśli chcesz x, zrób y”] okaże się prawdziwych w przypadku interpretacji niestandardowej [...], będziemy z pewnością odnosili sukcesy, z pewnością *przetrwamy* [...] i będziemy mieli potomstwo (Putnam 1981, s. 40).

Zarzut Douvena jest krótki i precyzyjny: „Załóżmy, że mówię językiem R, a mój sobowtór na Ziemi Bliźniaczej mówi językiem R\*. Oczywiście każde żywione przeze mnie prawdziwe/fałszywe przekonanie dyrektywne będzie prawdziwe/fałszywe zarówno w R, jak i R\*; tak więc mój sobowtór i ja będziemy mieli tę samą liczbę prawdziwych i fałszywych przekonań dyrektywnych. Ale, na przykład, przekonanie dyrektywne «Jeśli chcesz przetrwać, trzymaj się z daleka od tygrysów» może w R\* mieć treść następującą: «Jeśli pragniesz towarzystwa, kup sobie zwierzę domowe». Oba są prawdziwe, lecz trudno utrzymywać, że oba mają tę samą «wartość służącą przetrwaniu»”.

Zarzut ten jest zupełnie błędny, gdyż Douven (pisząc swój artykuł wcześniejszy) całkowicie pomylił ze sobą *odniesienie* i *wąską treść* (*narrow content*). Argument teoriomodelowy opiera się na założeniu (niekwestionowanym przez moich krytyków, w tym również przez Douvena), że każdy stan mentalny ma to, co Jaegwon Kim nazywał „rdzeniem wewnętrznym” (Kim 1993, s. 189-190). W przypadku percepcji rdzeniem wewnętrznym mają być dane zmysłowe posiadania tej percepcji (to, jakie muszą to być dane, jeśli teoria ma być „epistemicznie idealna”, jest wyznaczane przez „wymogi operacyjne”), natomiast w wypadku myśli dyrektywnej, takiej jak „Jeśli chcesz przetrwać, trzymaj się z daleka od tygrysów”, owym „trzonem wewnętrznym” byłoby żywienie myśli, której wąska treść kieruje nami następująco: „Jeśli chcemy przetrwać, działajmy tak, żeby nam się wydawało [żeby takie były nasze dane zmysłowe], iż trzymamy się z daleka od tygrysów”. Bardziej dokładnie, wszystkie myśli pozostają bez zmian zarówno na Ziemi, jak i Ziemi Bliźniaczej, zarówno pod względem syntak-

<sup>10</sup> *Could Reference Be Indeterminate?*, artykuł przedstawiony na zjeździe Wschodniego Oddziału Amerykańskiego Towarzystwa Filozoficznego w 1994 roku, lecz, o ile mi wiadomo, nieopublikowany [W korespondencji prywatnej Douven potwierdził, że faktycznie nie został on opublikowany, gdyż była to w zasadzie wstępna wersja artykułu późniejszego — przyp. tłum.].

tycznym, jak i każdym innym, włącznie z ich wąską treścią, lecz *to, do czego się odnoszą słowa posiadające te wąskie treści*, jest czymś odmiennym w R i R\*.

Douven zakładał, że R\* przypisuje nie tylko inną szeroką treść, ale także inną *wąską treść*. Ale to właśnie było błędne. A ponieważ jest to błędne, nic, co robię i co obserwuję nie może rozstrzygnąć, czy z Boskiego Punktu Widzenia „metafizycznie wyróżnioną” relacją odniesienia dla mojego języka jest R, czy też R\*.

Zrozumiawszy to, nie powinna, jak sądzę, zaskoczyć nas prawidłowa konkluzja, do jakiej Douven doszedł w swoim nowym artykule, a mianowicie, że twierdzenie, w myśl którego tylko jedna relacja odniesienia („związek przyczynowy”) jest jakoś metafizycznie wyróżniona, nie może być twierdzeniem *empirycznym*, teorią, która „musi być zaakceptowana lub odrzucona na podstawie dokładnie tych samych kryteriów, jakie decydują o wyborze teorii w obrębie tego, co tradycyjnie uważamy za nauki przyrodnicze”. Realści metafizyczni *muszą* pogwałcić swój własny naturalizm semantyczny. Zaczynam jednak wyprzedzać samego siebie.

Powróćmy teraz do późniejszego artykułu Douvena. Autor rozpoczyna od krótkiego i przejrzystego streszczenia argumentu teoriomodelowego:

Teoriomodelowy rdzeń argumentu Putnama jest w istocie całkiem prosty i jak najbardziej prawidłowy. Załóżmy, że  $T_1$  jest epistemicznie idealną teorią (tzn. teorią mającą każdą dającą się pomyśleć zaletę oprócz, być może, prawdziwości). Będąc idealną będzie ona między innymi niesprzeczna. Z pewnej elementarnej teorii modeli będzie wówczas wynikać, że  $T_1$  ma modele o każdej nieskończonej liczbie kardynalnej. Wybierzmy spośród nich jakiś model M o tej samej liczbie kardynalnej co świat i ustanówmy *relację spełniania* (SPEŁ) przez wzajemnie jednoznaczne odwzorowanie M na świat. Jeśli język teorii  $T_1$  zostanie zinterpretowany zgodnie z SPEŁ,  $T_1$  okaże się teorią prawdziwą o świecie, tzn.  $T_1$  będzie prawdziwa-według-SPEŁ, lub PRAWDZIWA(SPEŁ) (Douven 1999, s. 480-481).

Jednakże w następnym akapicie Douven popełnia spory błąd, pisząc: „Należy zauważyć, że, oprócz niesprzeczności, idealność  $T_1$  nie odgrywa w tym argumencie żadnej roli. Tak więc w istocie jest to argument dotyczący teorii niesprzecznych”. To dziwne stwierdzenie jest broniące w następujący sposób. Wpierw Douven przedstawia następujący zarzut w imieniu realisty metafizycznego:

Realista przyznaje, iż istnieje obfitość relacji korespondencji pomiędzy naszym słownikiem a rzeczywistością, lecz podkreśla, że tylko jedna z nich jest tą zamierzoną, tylko jedna z nich wyznacza odniesienie. O co zatem chodzi w tym argumencie? (Douven 1999, s. 481).

Kluczowa część odpowiedzi Douvena na to pytanie (na tym etapie jego wyводу — gdyż później daje inną, bardziej interesującą, odpowiedź) jest taka:

[...] Putnamowi chodzi o to, z grubsza rzecz biorąc, że dla realisty metafizycznego nie ma sposobu odróżnienia pomiędzy prawdą według jakiejś interpretacji a prawdą po prostu. [...] Realista ma tylko dwie możliwości nadania istotnej treści pojęciu zamierzoneści.<sup>11</sup> może zdefiniować „zamierzoneść” w odniesieniu do interpretacji, tak, że będzie ono oznaczało albo bycie

<sup>11</sup> „Zamierzoneść” odnosi się oczywiście do określenia „zamierzona” w wyrażeniu „zamierzona interpretacja”.



elementem klasy F-interpretacji (tzn. klasy interpretacji, które czynią teorię prawdziwą), albo też bycie elementem klasy nie-F-interpretacji. Ponieważ chcemy oczywiście by nasze teorie były prawdziwe, ta pierwsza możliwość wydaje się być lepszą drogą do definicji. Wówczas jednak okazuje się, że każda niesprzeczna teoria empiryczna musi być prawdziwa, wbrew FM [fallibilizmowi metodologicznemu] (Douven 1999, s. 481-482).

Jednakże twierdzenie, iż „każda niesprzeczna teoria empiryczna musi być prawdziwa” nie wchodziło w skład tego, za czym argumentowałem! Żeby zobaczyć, co jest nie tak z argumentem Douvena przedstawianym w moim imieniu, rozważmy „teorię” składającą się z jednego zdania „Właśnie widzę przed sobą krowę”.

To zdanie jest oczywiście niesprzeczne i wobec tego prawdziwe w świetle całego mnóstwa interpretacji (na przykład interpretacji, zgodnie z którą w moim przypadku znaczy ono „Mówię po angielsku”). Jednak (a w każdym razie w tej chwili) nie *wydaje mi się*, że bym widział krowę. Nie mam „krowich danych zmysłowych w moim polu widzenia”. I chociaż jest prawdą, że „chcę, by moje teorie były prawdziwe”, to chcę nie tylko tego. Nie zakładając czegoś, co najbardziej gorliwi antyrealiści lub najbardziej gorliwi realiści metafizyczni nie byliby w stanie zaakceptować, mogę powiedzieć, iż *chcę by moje teorie spełniały wymóg, że z reguły nie implikują one, iż widzę coś, kiedy mnie samemu nie wydaje się bym widział cokolwiek takiego*. Powiadam „z reguły”, ponieważ możemy dopuścić, bym czasem nie wiedział, co widzę. Jednakże akceptowanie teorii w całej rozciągłości tylko dlatego, że są one niesprzeczne, chociaż ich przewidywania obserwacyjne wydają się fałszywe, byłoby zupełnym porzuceniem metodologii empirycznej. To dlatego mój argument, z całym szacunkiem dla Douvena, nie jest „w istocie argumentem dotyczącym teorii niesprzecznych”, lecz co najmniej teorii, które są niesprzeczne *oraz* (w rozsądnym stopniu) spełniają nasze wymogi operacyjne, teorii implikujących przewidywania, które wydają się być prawdziwe.

Wszelako to nie wystarczy, ponieważ (przykładów jest cała rzesza) czasami odrzucamy teorię, która nie implikuje fałszywych przewidywań, lecz nie implikuje także żadnych prawdziwych przewidywań, które nie byłyby implikowane przez prostszą, bardziej elegancką lub silniejszą teorię z nią sprzeczną. Taka teoria gwałci to, co w *Realism and Reason* (1977) nazwałem „wymogiem teoretycznym”. Teoria, która nie narusza żadnego wymogu operacyjnego i teoretycznego może być nazwana „epistemicznie idealną”. Mój argument dotyczył właśnie teorii *epistemicznie idealnych*. Całkowicie błędne jest założenie Douvena, iż argumentowałem, że

nie ma żadnej cechy lub zbioru cech F\* innego niż F, różnicującego wszystkie istniejące relacje korespondencji pomiędzy słowami języka teorii a rzeczami lub zbiorami rzeczy w świecie (Douven 1999, 482).

Mój argument polegał na tym, że aczkolwiek niektóre relacje korespondencji *mogą* być wykluczone — gdyż teorie czynione przez nie prawdziwymi okazują się *epistemicznie* nieakceptowalne, chociaż są niesprzeczne — to realista metafizyczny nie może wskazać *nie-epistemicznej* cechy różnicującej wszystkie istniejące relacje kore-

spondencji pomiędzy słowami języka teorii a rzeczami lub zbiorami rzeczy w świecie, która — by użyć kolejnego zwrotu Douvena — „spełniałaby standardy realisty metafizycznego”.

Na szczęście argument, który Douven podaje (w moim imieniu) na poparcie tego ostatniego twierdzenia *mogę* zaakceptować, pomimo błędu, który właśnie wskazałem (przy założeniu, że cechy *epistemiczne*  $F^*$  są irrelewantne dla mojego argumentu). Twierdzenie to (jest to jego twierdzenie („3”)) formułuje on następująco: „(3) Nie ma takiej cechy  $F^*$ , która spełniałaby standardy realisty metafizycznego” (Douven 1999, s. 482).

Podążajmy dalej za rekonstrukcją Douvena. Piszę on:

(3) zdumiało wielu realistycznych oponentów Putnama. Typowo naturalistycznie nastawieni realisci przyjmują jakąś wersję *przyczynowej teorii odniesienia* (PTO), w myśl której odniesienie jest konstytuowane przyczynowo.<sup>12</sup> Innymi słowy, twierdzą oni, iż istnieje akceptowalna własność  $F^*$  — przyczynowość — wystarczająca do wyodrębnienia odniesienia z gąszczy relacji słowo/świat. Odpowiedź realisty na argument teoriomodelowy Putnama jest zatem najwyraźniej taka: ponieważ nie ma gwarancji, iż SPEL pokrywa się z relacjami przyczynowymi konstytuującymi odniesienie, nie ma gwarancji, że  $T_1$  [epistemicznie idealna teoria] jest prawdziwa (zamiast być po prostu PRAWDZIWA(SPEL)) (Douven 1999, s. 483).

Jak wyjaśnia Douven, antycypowałem tę odpowiedź i szybko się z nią uporałem (zapewne zbyt krótko, jak widać z perspektywy czasu), pisząc:

Zauważmy, iż „przyczynowa” teoria odniesienia nie jest tu (nie byłaby) jakkolwiek pomocą, ponieważ w obrazie realisty metafizycznego to, jak może się jednoznacznie odnosić wyrażenie „wywołuje przyczynowo” jest czymś tak samo zagadkowym jak to, jak może odnosić się słowo „kot” (Putnam 1977, s. 486; przedruk, s. 126).

Douven pisze, że ta uwaga jest zazwyczaj interpretowana w sposób następujący: przyjąwszy, iż PTO jest rzeczywiście teorią empiryczną (jak to jest wymagane przez naturalizm semantyczny) i że nie narusza ona jakichkolwiek wymogów empirycznych (ponieważ gdyby naruszała, podlegałaby empirycznej lub metodologicznej dyskonfirmacji), PTO jest częścią całościowej epistemicznie idealnej teorii.

Jednakże teraz argument teoriomodelowy może, z uwagi na jego ogólność, zostać powtórzony [dla całościowej teorii, włącznie z PTO]. Jest zupełnie oczywiste — zgodnie z kolejnym etapem tak odczytanego argumentu — że jeśli założymy, iż problem, z którym boryka się realista został już rozwiązany dla tej części języka, w którym sformułowana jest PTO, to wtedy ta teoria może być prawdopodobnie przyjęta dla rozwiązania owego problemu dla  $T_1$  (tak samo jak dla każdej innej teorii, do której stosuje się argument teoriomodelowy), lecz „w obrazie realisty metafizycznego to, jak może się jednoznacznie odnosić wyrażenie «wywołuje przyczynowo»

<sup>12</sup> Krytykę koncepcji „przyczynowości”, zakładanej przez takie teorie, zawierają moje rozważania w książce *Renewing Philosophy* na temat Fedorowskiej wersji teorii przyczynowej (Putnam 1992). Fakt, że to, do czego odnosimy się jako do „przyczyny” danego zdarzenia, zależy od naszych *zainteresowań*, a wobec tego nie jest czymś uprzednio ustalonym i niezależnym od naszej intencjonalności, jest zupełnie ignorowany przez obrońców tzw. przyczynowej teorii odniesienia!

jest czymś tak samo zagadkowym, jak to jak może odnosić się słowo «kot»». Innymi słowy, PTO jest częścią problemu, a nie poszukiwanym rozwiązaniem. Przyjmować inaczej to z góry przesądzać sprawę (Douven 1999, s. 483-484).

Tak odczytany argument był szeroko oskarżany o to, że jest „nieprawidłowy i niedopuszczalny” (Hale, Wright 1997, s. 440 nn.) i że „po prostu z góry przesądza sprawę na niekorzyść realisty” (Devitt 1991, s. 227). Jednak, jak wskazuje Douven, przy takim odczytaniu fakt, iż przyjmuje się, że PTO jest *teorią empiryczną*, nie odgrywa żadnej istotnej roli. Jeśli do idealnej teorii empirycznej dołączymy czysto metafizyczną (to znaczy empirycznie nietestowalną) PTO, albo jakąkolwiek inną nietestowalną teorię odniesienia, argument ten będzie funkcjonował równie dobrze.

W myśl rekonstrukcji Douvena zmuszam realistę metafizycznego do stawienia czoła problemowi, który istotnie zależy od jego ambicji bycia naturalistą w semantyce (pisząc *Reason, Truth and History* twierdziłem, że w tej postaci problem nie powstaje dla czegoś, co nazwałem „magiczną teorią odniesienia” — co nie jest, oczywiście, argumentem za przyjęciem teorii magicznej!). Niezwykle pomysłowy sposób, w jaki Douven to rozwija, zależy od jego błędnego poglądu, że argument teoriomodelowy może być zastosowany do każdej niesprzecznej teorii, ale jego pomysł daje się łatwo uwolnić od tego defektu. Oto moja własna naprawiona wersja argumentu Douvena:

Załóżmy, że PTO jest faktycznie teorią empiryczną. Są dwie możliwości: albo (1) PTO wynika dedukcyjnie z *niesemantycznej* części idealnej teorii  $T_1$ , albo (2) PTO jest logicznie niezależna od niesemantycznej części teorii  $T_1$ . Jeśli przyjmiemy (1), to realista metafizyczny jest nam winien wyjaśnienie, jak teoria semantyczna może wynikać z niesemantycznych faktów empirycznych. (Musiałby on podać *analityczną* definicję „odniesienia” w kategoriach niesemantycznych!) O ile wiem, żaden realista metafizyczny nie przyjmuje, iż jest to możliwe. Pozostaje nam zatem (2).

Ponieważ PTO jest teorią o treści empirycznej, niezależną od niesemantycznej części nauki idealnej, istnieje logicznie możliwy świat, w którym niesemantyczna część nauki idealnej jest prawdziwa, natomiast PTO jest fałszywa<sup>13</sup>. Jeśli taki świat jest aktualny, to — ponieważ nikt nie sugerował żadnego *innego* naturalistycznego wymogu, nakładanego na zbiór dopuszczalnych relacji odniesienia, niż „relacje przyczynowe”, a relacje przyczynowe nie wyodrębniają uprzywilejowanej relacji odniesienia, jeśli PTO jest fałszywa — w takim świecie (jak argumentuje Douven) argument teoriomodelowy jest poprawny: każda relacja SPEŁ, która czyni idealną teorię prawdziwą, jest dopuszczalną relacją odniesienia, a żadna z wielu dopuszczalnych relacji odniesienia nie jest uprzywilejowana. Ponieważ *cała* teoria idealna, *łącznie ze słowami wyrażającymi PTO*, jest prawdziwa w wypadku tych wszystkich relacji SPEŁ, to wynika z tego, że w świecie, który właśnie sobie wyobraziliśmy,

<sup>13</sup> Realistom metafizycznemu nie pomoże twierdzenie, że nawet gdyby taki świat był logicznie możliwy, to nie jest on „metafizycznie możliwy”. To pojęcie jest bowiem, jak sądzę, beznadziejnie problematyczne. (W sprawie szczegółowej krytyki zob. Putnam 1990).

słowa użyte przez nas do wyrażenia PTO są prawdziwe (prawdziwe w ramach wszystkich dopuszczalnych relacji SPEŁ), chociaż PTO nie jest „w rzeczywistości” prawdziwa. Okazuje się zatem, że nawet w światach, w których PTO jest fałszywa, musimy i tak zaakceptować słowa wyrażające PTO jako prawdziwą, o ile tylko niesemantyczna część teorii idealnej nie jest empirycznie obalona! Krótko mówiąc (przyjmując, że niesemantyczna część naszej teorii jest ustalona), PTO gwałci FM (fallibilizm metodologiczny). W pewnym sensie sytuacja jest podobna do tej, którą przedstawiłem w moim argumencie z mózgów w naczyniu w rozdziale pierwszym *Reason, Truth and History* (Putnam 1981): słowa, które wyrażają PTO są semantycznie niestabilne, gdyż to, do czego się odnoszą, zależy od świata, w którym się znajdujemy. Wszelako odmiennie od słów „Jesteśmy mózgami w naczyniu”, które zmieniły odniesienie w taki sposób, że okazywały się zawsze fałszywe, niezależnie od świata, w którym była osoba je wypowiadająca, PTO zmienia warunki odniesienia w taki sposób, iż pozostaje prawdziwa *nawet wówczas, gdy realizm metafizyczny jest fałszywy*.

Jest to niezwykle pomysłowy argument. Nie zamierzam jednak twierdzić, iż był to *mój* argument w *Realism and Reason*. Przedstawiłem tam znacznie prostszy argument, prowadzący do tego samego wniosku, a mianowicie, że pytanie o to, która relacja odniesienia jest „zamierzona” lub „metafizycznie wyróżniona”, a także pytanie o to, czy dana jednoznaczna relacja odniesienia jest „zamierzona” lub „metafizycznie wyróżniona”, nie są pytaniami *empirycznymi* i dlatego realista metafizyczny *nie może* być jednocześnie naturalistą semantycznym. To ten prostszy argument Douven błędnie odrzucił w swoim wcześniejszym artykule. Niezależnie od tego, czy ktoś jest przekonany, że istnieje jedna zamierzona relacja odniesienia, czy też, że wszystkie relacje SPEŁ czyniące  $T_1$  prawdziwą są równie dopuszczalne i żadna z nich nie jest ową jedynie uprzywilejowaną, to — jak twierdziłem w *Realism and Reason* i szczegółowo pokazałem w *Reason, Truth and History* — nie ma to jakiegokolwiek znaczenia dla przewidywań, które na temat czegokolwiek będą formułowane. Te odmiennie poglądy metafizyczne są *empirycznie nieodróżnialne*. Oznacza to, że PTO jest „semantyką”, która nie spełnia maksymy, iż „semantyka jest nauką empiryczną, taką samą jak każda inna”.

Oczywiście realista metafizyczny może odpowiedzieć na argument Douvena twierdząc, iż nawet jeśli PTO jest nierewidowalna, pod warunkiem, że niesemantyczną część naszej teorii świata T przyjmuje się za ustaloną, to i tak podlega ona rewizji w szerszym sensie, gdyż istnieją pewne części T — pewne niesemantyczne przekonania, które teraz posiadamy i które uważamy także za część *idealnej* teorii  $T_1$  — takie, iż ich empiryczne obalenie prowadziłyby do porzucenia PTO. Realista metafizyczny może rozsądnie twierdzić, że wymaganie, by PTO była fałszyfikowalna *nawet przy przyjęciu niesemantycznej części  $T_1$  jako ustalonej*, jest niedorzeczną interpretacją fallibilizmu metodologicznego (ponieważ w rzeczywistości wymaga ona, by każda logicznie niezależna część teorii empirycznej była niezależnie *testowalna*). Wystarczy, jeśli istnieją dające się przedstawić odkrycia empiryczne, które wyma-

gałyby od nas porzucenia PTO, chociaż wymagałyby one od nas także zmiany nie-semantycznej części  $T_1$ .

Jeśli realista metafizyczny obierze tę drogę, to zmuszony jest twierdzić, iż istnieją możliwe *odkrycia niesemantyczne* obalające PTO. Jednak, o ile mi wiadomo, żaden realista metafizyczny czegoś takiego nie sugerował ani nie sygnalizował, jak taki eksperyment miałby wyglądać!

I jeszcze jedno spostrzeżenie. Filozof, który akceptuje argument teoriomodelowy i tym samym jest przekonany, iż nie istnieje jedyna „zamierzona” relacja odniesienia, może jeszcze przyjąć metametajęzyk MML, w którym powie on o danym języku przedmiotowym L, że pewna interpretacja L (w metajęzyku ML) jest „zamierzoną interpretacją L”. (Może on nawet uznać PTO). Wtedy jednak w pewnym innym języku MMML stwierdzi on, że sam metametajęzyk MML ma „więcej niż jedną zamierzoną interpretację”. Jak napisałem w *Realism and Reason*: „pytanie o to, czy teoria ma *jedyną* interpretację *nie ma* absolutnego sensu” (Putnam 1977, s. 494; przedruk, s. 136). „Posiadanie jedynej zamierzonej interpretacji” samo okazuje się *własnością zrelatywizowaną do teorii*. (Przy założeniu, którego ja sam już nie przyjmuję, że prawomocna jest konkluzja argumentu teoriomodelowego).

#### UWAGI KOŃCOWE

Podkreślałem, że jesteśmy zmuszeni przyjąć argument teoriomodelowy, a wraz z nim weryfikacjonistyczny pogląd, iż epistemicznie idealna teoria nie może nie być prawdziwa tylko wtedy, gdy przyjmiemy *założenia*, na których argument się opiera. Owe założenia zawierają nie tylko ideę „wymogów operacyjnych”, z jej ukrytym odwołaniem się do obrazu danych zmysłowych, lecz także to, co sam określiłem mianem „weryfikacjonistycznej koncepcji rozumienia”. Czy to zaskakujące, że przyjmąwszy tak wiele z pozytywizmu logicznego, jest się później zmuszonym do przyjęcia pozytywistycznych konkluzji? Podejrzewam, iż ci sami krytycy, którzy uważali konkluzję argumentu teoriomodelowego za kapitulację przed pozytywizmem, nie zdolali poddać krytyce pozytywistycznych *założeń* argumentu, ponieważ założenia te *akceptują* — w istocie rzeczy owe założenia wydają się być im narzucane przez kartezyjanizm powiązany z materializmem, który w dzisiejszych czasach tak często uchodzi za „kognitywistykę”. A przecież to właśnie w *tym punkcie* — co pierwszy zauważył Cesar Gomez — argument teoriomodelowy powinien być krytykowany.

W moich wykładach im. Deweya (Putnam 1994a) przedstawiłem sposób, w jaki można uniknąć całego tego argumentu przez porzucenie „interfejsowej koncepcji percepcji” i „interfejsowej koncepcji konceptualizacji”, przy których trwamy od trzech stuleci. Stwierdzenie tego nie ma wymowy „antynaukowej”. Jak napisałem w wykładach im. Deweya:

W naszym realizmie zdroworozsądkowym dotyczącym zarówno percepcji, jak i konceptualizacji, nie ma nic, co byłoby „nienaukowe” w sensie *piętrzenia przeszkód* przed poważnymi pró-

bami dostarczenia lepszych modeli, zarówno neurologicznych, jak i obliczeniowych, procesów mózgowych, od których zależą nasze zdolności percepcyjne i konceptualne, procesów, o których nadal wiemy tak mało. Co więcej, jest ogromnym błędem zrównywanie prawdziwej nauki z kartezjanizmem powiązany z materializmem, który od trzech stuleci stara się występować pod płaszczykiem nauki (Putnam 1994a, s. 494; przedruk, s. 48).

#### APENDYKS

#### CO ROZUMIEM PRZEZ „REALIZM ZDROWOROZSĄDKOWY”<sup>14</sup>

Wielu filozofów przyznało, że są zdezorientowani w kwestii tego, co właściwie miałem na myśli mówiąc o powrocie do „realizmu zdroworozsądkowego” w wykładach im. Deweya. W skrócie, „realizm zdroworozsądkowy”, w moim sensie tego słowa, zawiera w sobie element negatywny, a mianowicie, odrzucenie poglądu, że prawda nie może przekraczać weryfikowalności,<sup>15</sup> oraz dwa elementy pozytywne: powrót (tak bardzo, jak to możliwe) do „realizmu naiwnego” w odniesieniu do percepcji i odcudzysłowieniową (*disquotational*) koncepcję prawdy, podobną do tej, jaką dostrzegam u Wittgensteina. Różni się on od tego, co kiedyś *nazwałem* „realizmem metafizycznym”, gdyż odrzuca to, co uważam za fantazjowanie na temat jednej ostatecznie prawdziwej i zupełnej Ontologii, lecz, oczywiście, *jest* na swój sposób zarazem metafizyczny i realistyczny. Oto krótkie objaśnienie tych dwóch elementów pozytywnych.

#### Realizm naiwny

W mojej książce *The Threefold Cord: Mind, Body, and World* (Putnam 1999), częściowo pod wpływem poglądów Johna McDowella z *Mind and World* (McDowell 1994), skoncentrowałem się na potrzebie uniknięcia myślenia czy to o doświadczeniach percepcyjnych, czy to o myślach, jako „ekranach” *pomiędzy* nami a światem. Dzisiaj część stanowiska, za którym wtedy argumentowałem, tj. twierdzenie, że podczas normalnej weredyckiej percepcji obiektów w naszym otoczeniu (np. wiązanki kilku czerwonych róż) przedmiotem naszej świadomości nie są obiekty mentalne lub qualia, lecz same te obiekty zewnętrzne (czerwone róże), stała się ortodoksją wśród filozofów umysłu. Obraz naszych umysłów jako spoglądających na „wewnętrzny ekran filmowy” jest zdecydowanie niemodny. Niemal wszyscy filozofowie percepcji chcą powrócić, przynajmniej częściowo, do „realizmu naiwnego” (termin, który obecnie wydaje się być modny na oznaczenie tego, co we wspomnianej książce *nazwałem* „realizmem naturalnym”). Jednakże powrót do realizmu naturalnego jest bardziej skomplikowany niż to przedstawiłem w *The Threefold Cord*. Jedną z oznak

<sup>14</sup> Czynień tu użytek z dwóch lub trzech akapitów z mojej autobiografii intelektualnej, zawartej w: *The Philosophy of Hilary Putnam*, ed. R.E. Auxier, Chicago, Open Court (w przygotowaniu).

<sup>15</sup> Na ten temat zob. moją wymianę z Crispinem Wrightem (Wright 2000 i Putnam 2001).

złożoności zagadnień z tym związanych jest to, iż obok stanowisk, które omawiałem w moich wykładach im. Deweya i im. Royce'a (w dwóch seriach wykładów, które potem zostały opublikowane w książce *The Threefold Cord*), tj. tradycyjnej teorii danych zmysłowych i „alternatywizmu” (np. Johna McDowella), istnieją obecnie także „fenomeniści” (np. Ned Block) i „reprezentacjoniści” (np. Michael Tye), i każda z tych szkół ma swoje frakcje, które różnią się w ważnych kwestiach. W dodatku większość (choć nie wszyscy) z tych filozofów, włączając w to mnie samego, jest „szerokimi funkcjonalistami”, tzn. pojmują oni nasze udane percepcje jako rezultat obejmujących świat (*world-involving*) stanów funkcjonalnych, chociaż na ogół nie stanów *obliczeniowych*. (Ten ostatni fakt jest oczywiście związany z porzuceniem obrazu percepcji, polegającej na skanowaniu wewnętrznego „ekranu filmowego” oraz z faktem, że wszyscy ci filozofowie chcą oddać sprawiedliwość temu, co jest słuszne w „realizmie naiwnym”).

### Dyskwotacjonizm

Wszelako jest jeszcze inny element mojego „realizmu zdroworozsądkowego”. Drugim elementem na drodze do filozoficznej obrony „realizmu naturalnego”, którą przedstawiłem w wykładach im. Deweya, jest „odczuysłowieniowa koncepcja prawdy”. Aby to wyjaśnić, krótko opiszę pewną rodzinę (dwudziestowiecznych) teorii prawdy. Rodzina ta jest powszechnie określana za pomocą terminu, którego właśnie użyłem: „odczuysłowieniowe” koncepcje prawdy. W wykładach im. Deweya argumentowałem, iż są co najmniej trzy bardzo odmienne wersje „odczuysłowienia”: wersja Fregego; wersja „deflacyjna”, reprezentowana przez Carnapa (1949), i wersja broniona przeze mnie (i przypisywana późnemu Wittgensteinowi).

Tym, co łączy te koncepcje jest nacisk, jaki kładą one na to, co Michael Dummett nazywa Fregowską „zasadą równoważności” (jest to oczywista antycypacja „konwencji T” Tarskiego) (Dummett 1973 i 1981; Tarski 1956), która głosi, że stwierdzać lub sądzić, iż jest czymś prawdziwym, że śnieg jest biały, lub jest czymś prawdziwym, że morderstwo jest czymś złym, lub jest czymś prawdziwym, że dwa jest jedyną parzystą liczbą pierwszą (lub jaki nie byłby to przykład), jest czymś równoważnym sądzeniu, że — odpowiednio — śnieg jest biały, lub morderstwo jest czymś złym, lub dwa jest jedyną parzystą liczbą pierwszą. Jeśli przyjmie się, że sądzenie, iż jest czymś prawdziwym, że śnieg jest biały (*etc.*), jest tym samym, co sądzenie, że *zdanie* (w tym, co Tarski nazwał „językiem przedmiotowym”) „Śnieg jest biały” jest prawdziwe (jest to jedna z tych kwestii, która odróżnia Tarskiego od Fregego), wówczas można wyrazić zasadę równoważności poprzez napisanie (zgodnie z Tarskim):

(T) „Śnieg jest biały” jest prawdziwe wtedy i tylko wtedy, gdy śnieg jest biały

Kiedy mówię, że wszystkie wersje dyskwotacjonizmu „kładą nacisk” na jakąś wersję zasady równoważności, mam na myśli rzeczy następujące:

(1) Dyskwotacjoniści utrzymują, że kiedy słowo „prawdziwy” jest stosowane do stwierdzenia  $S$ , które jest wprost dane, to jest ono *eliminowalne*. Na przykład w niektórych miejscach Frege wydaje się utrzymywać, że metastwierdzenie, takie jak „Jest prawdziwe, że śnieg jest biały”, i odpowiadające mu stwierdzenie przedmiotowe „Śnieg jest biały”, wyrażają jeden i ten sam sąd. Gdybyśmy używali słowa „prawdziwy” tylko w zdaniach o postaci: „ $S$  jest prawdziwe”, gdzie  $S$  jest cytatem zdania, lub w zdaniach o postaci „Jest prawdziwe, że  $p$ ”, gdzie  $p$  jest zdaniem, to wtedy słowo to byłoby *niepotrzebne*.

(2) Według dyskwotacjonistów, powodem, dla którego potrzebujemy słowa „prawdziwy” i jego synonimów w naszych językach, jest to, że musimy być zdolni do używania słowa „prawdziwy” w zdaniach o postaci „ $x$  jest prawdziwe”, gdzie „ $x$ ” jest *zmienną kwantyfikacji*, a nie — powiedzmy — zdaniem w cudzysłowie. Dla przykładu, jeśli powiem: „Przynajmniej jedno ze zdań, które Jan napisał na stronie 12 jest prawdziwe” [w notacji rachunku predykatów:  $(\exists x) (x \text{ jest napisane przez Jana na stronie 12 i } x \text{ jest prawdziwe)}$ ], to mogę „wylimitować” słowo „prawdziwy” *wtedy i tylko wtedy*, gdy wiem, jakie zdania Jan napisał na stronie 12. Np. jeśli wiem, że jedynymi zdaniami, jakie Jan napisał na stronie 12 są: „Śnieg jest biały”, „Morderstwo jest złem” i „Dwa jest jedyną parzystą liczbą pierwszą”, to wtedy wiem, iż to stwierdzenie ma taką samą wartość logiczną jak: „Jan napisał «Śnieg jest biały» na stronie 12 i śnieg jest biały, lub Jan napisał «Morderstwo jest złem» na stronie 12 i morderstwo jest złem, lub Jan napisał «Dwa jest jedyną parzystą liczbą pierwszą» na stronie 12 i dwa jest jedyną parzystą liczbą pierwszą, i są to jedyne zdania jakie Jan napisał na stronie 12”, gdzie owo przydługie zdanie nie zawiera słowa „prawdziwy”. Jeśli jednak nie wiem, *jakie* zdania Jan napisał, to nie wiedziałbym, jak skonstruować materialnie równoważne zdanie niezawierające słowa „prawdziwy”.

Podsumowując: wszyscy dyskwotacjoniści zgadzają się, że predykat z logiczną własnością prawdziwości („własnością odcudzysłowienia”) jest konieczny z powodów *logicznych*, a nie *opisowych*.

Różnice pomiędzy różnymi wersjami odcudzysłowienia są, z grubsza rzecz biorąc, następujące. Dla Fregego słowo „prawdziwy” jest predykatem orzekanym o „myślach” (*Gedanken*), a myśli nie składają się ze słów. Czy są one obiektami platońskimi, czy też wykraczają poza dychotomię platońskie — nieplatońskie, jest przedmiotem sporów wśród badaczy Fregego; w każdym razie istnieje metafizyka „myśli” presuponowana przez wersję Fregego. Co więcej, myśli o realnościach empirycznych są o nich *ze swej wewnętrznej* natury; dla Fregego myśli nie są obrazami malowanymi „mentalną farbą”, które w jakiś sposób „korespondują” ze światem; *stwierdzają* one coś o świecie. Dla Tarskiego, wprost przeciwnie, słowo „prawdziwy” jest predykatem orzekanym o zdaniach, a *zdania są zwykłymi ciągami znaków na papierze*. (Tarski co prawda mówi, że przyjmuje, iż owe znaki posiadają „konkretne znaczenia”, lecz ten wymóg nie jest częścią *definicji* prawdy, co czyni prawdę po prostu własnością znaków, nie zaś własnością, którą mają wtedy, *kiedy* posiadają owe (nie-



zdefiniowane) „konkretne znaczenia”<sup>16</sup>). Sformułowane przez Tarskiego „definicje prawdy”, wbrew temu, co się często głosi, ani nie zakładają, ani nie prowadzą do „korespondencyjnego” pojęcia prawdy. Dla Carnapa, który w całej pełni akceptował teorię Tarskiego, rozumiemy zdanie, powiedzmy „Śnieg jest biały”, rozumiejąc procedurę jego weryfikacji, i rozumiemy odpowiadające mu metazdanie „Zdanie «śnieg jest biały» jest prawdziwe”, wiedząc, iż ma ono dokładnie taką samą procedurę weryfikacji. (To właśnie tę koncepcję „deflacyjną” odrzuciłem w wykładach im. Deweya). Dla Wittgensteina (i mnie samego) „prawdziwy” jest predykatem orzekanym o zdaniach używanych na pewne sposoby — tzn. orzekanym o obiektach, które nie są ani czysto syntaktyczne (jak „zdania” Tarskiego), ani nie są niezależne od obejmujących świat użyć obiektów syntaktycznych w jakiejś konkretnej społeczności językowej. Podobnie jak w koncepcji Fregego, prowadzi to do tego, że stwierdzenie takie, jak „Śnieg jest biały”, nie jest obrazem namalowanym mentalną farbą, który jakoś koresponduje z tym, że śnieg jest biały; użycie tego zdania *obejmuje* śnieg i białość. Jeśli istnieje tu „relacja korespondencji”, to jest ona relacją wewnętrzną, nie zewnętrzną lub przygodną.

Przełożyli Krzysztof Czerniawski i Tadeusz Szubka

## BIBLIOGRAFIA

- Carnap R. (1949), *Truth and Confirmation*, [w:] *Readings in Philosophical Analysis*, ed. H. Feigl, W. Sellars, New York, Appleton-Century-Crofts, s. 119-127.
- Devitt M. (1991), *Realism and Truth*, 2<sup>nd</sup> ed., Oxford, Blackwell 1991.
- Devitt M. (1996), *Coming to Our Senses. A Naturalistic Program for Semantic Localism*, Cambridge, Cambridge University Press.
- Douven I. (1999), *Model-Theoretic Argument Reconstructed*, „The Journal of Philosophy”, XCVI, no. 9, s. 479-490.
- Dummett M. (1973), *Frege: Philosophy of Language*, London, Duckworth.
- Dummett M. (1981), *The Interpretation of Frege's Philosophy*, Cambridge, MA, Harvard University Press.
- Hale B., Wright C. (1997), *Putnam's Model-Theoretic Argument Against Metaphysical Realism*, [w:] *A Companion to the Philosophy of Language*, ed. B. Hale, C. Wright, Oxford, Blackwell, s. 427-457.

<sup>16</sup> Fragment z rozprawy Tarskiego, do którego odwołuje się Putnam brzmi: „Zbyteczne jest może dodawać, że nie interesują tu nas wcale języki i nauki «formalne» w pewnym specyficznym znaczeniu tego wyrazu, a mianowicie tego rodzaju nauki, iż występującym w nich znakom i wyrażeniom nie przypisuje się żadnego intuicyjnego sensu; w odniesieniu do takich nauk postawione tu zagadnienie traci wszelką rację bytu i przestaje być po prostu zrozumiałe. Znakom, występującym w tych językach, których dotyczą niniejsze rozważania, przypisujemy zawsze całkiem konkretne i zrozumiałe dla nas znaczenie” (Tarski 1956, s. 166-167; oryginał polski, s. 17; przedruk, s. 33) [przyp. tłum.].

- Kim J. (1993), *Supervenience and Mind. Selected Philosophical Essays*, Cambridge, Cambridge University Press.
- McDowell J. (1994), *Mind and World*, Cambridge, MA, Harvard University Press.
- Putnam H. (1977), *Realism and Reason*, „Proceedings and Addresses of the American Philosophical Association”, L, no. 6, s. 483-498; przedruk w: Putnam 1978, s. 123-140.
- Putnam H. (1978), *Meaning and the Moral Sciences*, Routledge & Kegan Paul, London.
- Putnam H. (1980), *Models and Reality*, „The Journal of Symbolic Logic”, XLV, no. 3, s. 464-482 [Modele i rzeczywistość, [w:] tenże, *Wiele twarzy realizmu i inne eseje*, przeł. A. Grobler, Warszawa 1998, Wyd. Naukowe PWN, s. 185-224].
- Putnam H. (1981), *Reason, Truth and History*, Cambridge, Cambridge University Press [rozdział 1 dostępny po polsku: *Mózgi w naczyniu*, [w:] tenże, *Wiele twarzy realizmu i inne eseje*, przeł. A. Grobler, Warszawa 1998, Wyd. Naukowe PWN, s. 295-324].
- Putnam H. (1987), *The Many Faces of Realism*, LaSalle, IL, Open Court [*Wiele twarzy realizmu*, [w:] tenże, *Wiele twarzy realizmu i inne eseje*, przeł. A. Grobler, Warszawa 1998, Wyd. Naukowe PWN, s. 325-429].
- Putnam H. (1988), *Representation and Reality*, Cambridge, MA, A Bradford Book/The MIT Press.
- Putnam H. (1990), *Is Water Necessarily H<sub>2</sub>O*, [w:] tenże, *Realism with a Human Face*, ed. J. Conant, Cambridge, MA, Harvard University Press, s. 54-79.
- Putnam H. (1992), *Renewing Philosophy*, Cambridge, MA, Harvard University Press.
- Putnam H. (1994a), *Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind*, „The Journal of Philosophy”, XCI, no. 9, s. 445-517; przedruk w: Putnam 1999, s. 1-70.
- Putnam H. (1994b), *Simon Blackburn on Internal Realism*, [w:] *Reading Putnam*, ed. P. Clark, B. Hale, Oxford, Blackwell, s. 242-254.
- Putnam H. (1996), *Pragmatism and Realism*, „Cardozo Law Review”, XVIII, no. 1, s. 153-170.
- Putnam (1999), *The Threefold Cord: Mind, Body and World*, New York, Columbia University Press.
- Putnam (2001), *When „Evidence Transcendence” Is Not Malign: A Reply to Crispin Wright*, „The Journal of Philosophy”, XCVIII, no. 11, s. 594-600.
- Tarski A. (1956), *The Concept of Truth in Formalized Languages*, [w:] tenże, *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*, Oxford, Clarendon Press, s. 152-278 [*Pojęcie prawdy w językach nauk dedukcyjnych*, Warszawa 1933, Towarzystwo Naukowe Warszawskie; przedruk w: tenże, *Pisma logiczno-filozoficzne, t. 1: Prawda*, Warszawa 1995, Wyd. Naukowe PWN, s. 13-172].
- Wright C. (1992), *Truth and Objectivity*, Cambridge, MA, Harvard University Press.
- Wright C. (2000), *Truth as Sort of Epistemic: Putnam's Peregrinations*, „The Journal of Philosophy”, XCVII, no. 6, s. 335-364.