

Witold Strawiński

Funkcjonalistyczna koncepcja umysłu a postulat jedności nauki

Filozofia Nauki 12/3/4, 107-114

2004

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

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Witold Strawiński

Funkcjonalistyczna koncepcja umysłu a postulat jedności nauki¹

Funkcjonalizm, obok behawioryzmu i radykalnego fizykalizmu (teorii identyczności), stanowi jedno z najbardziej znanych ujęć zagadnienia psychofizycznego w filozofii analitycznej XX wieku. Głosi on, że stany umysłowe są ukonstituowane przez swoje wzajemne związki przyczynowe oraz przez przyczynowe powiązania ze zmysłowymi bodźcami (stanami „wejściowymi” — *inputs*) i behawioralnymi reakcjami (stanami „wyjściowymi” — *outputs*).² Niektórzy dostrzegają w tym stanowisku przede wszystkim kontynuację wątków behawiorystycznych.³ Ja będę chciał zwrócić uwagę na pewien inny aspekt kształtowania się i oddziaływania funkcjonalistycznego ujęcia umysłu. Chodzi o **antyredukcjonistyczną argumentację** kierowaną przeciwko skrajnemu fizykalizmowi Smarta i Feigl’a (tzn. przeciwko teorii identyczności umysłu i mózgu), m.in. przez D. Davidsona (monizm anomalistyczny)⁴ oraz przez H. Putnama (wielorakie realizacje stanów umysłowych)⁵, uogól-

¹ W niniejszym referacie wykorzystane zostały pewne fragmenty rozdz. 9 mojej książki: W. Strawiński, *Jedność nauki, redukcja, emergencja. Z metodologicznych i ontologicznych problemów integracji wiedzy*, Warszawa 1997, Aletheia.

² N. Block, *Functionalism*, [w:] J. Kim, E. Sosa, (red.), *A Companion to Metaphysics*, Oxford 1995, Blackwell., s. 189.

³ Por. np. S. Shoemaker, *Functionalism and Qualia*, „Philosophical Studies”, R. 27 (1975), s. 306-307.

⁴ D. Davidson, *Eseje o prawdzie, języku i umyśle*, Warszawa 1992, PWN, s. 163-193: *Zdarzenia mentalne*.

⁵ H. Putnam, *The Mental Life of Some Machines* (1967), *The Nature of Mental States* (1967); przedruk [w:] tenże, *Mind, Language and Reality. Philosophical Papers*, v. 2, Cambridge 1975: Cambridge University Press, s. 408-440.

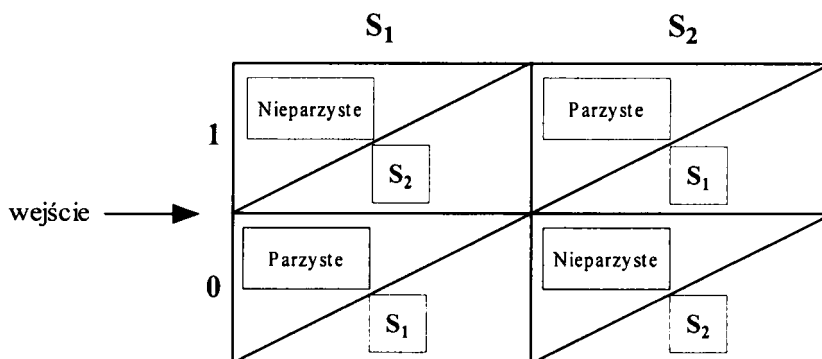
nioną i zwróconą w 1974 r. przeciwko logiczno-empirystycznemu programowi „jedności nauki” przez J. Fodora.⁶

Na początku przedstawię główne idee funkcjonalistycznej koncepcji umysłu, wykorzystując charakterystykę tej koncepcji przedstawioną przez Neda Blocka.⁷ Załóżmy, że dysponujemy pewną teorią stanów umysłowych (mentalnych), pozwalającą określić wszystkie związki przyczynowe między tymi stanami i odpowiednimi bodźcami sensorycznymi oraz reakcjami behawioralnymi. Na przykład, jeżeli chodzi o ból, ujmowany jako pewien umysłowy stan wewnętrzny, to teoria ta mogłaby stwierdzać, że ukłucie szpilką powoduje ból, a ból powoduje rozdrażnienie (kolejny stan wewnętrzny) i wypowiedzenie „Oj!”. Gdyby to było wszystko, co dana teoria ma do stwierdzenia na temat bólu, to moglibyśmy na jej gruncie zdefiniować ból następująco:

x doznaje bólu $\Leftrightarrow \exists P \exists Q$ (ukłucie szpilką przyczynowo powoduje stan wewnętrzny P i P przyczynowo powoduje zarazem stan wewnętrzny Q oraz wypowiedzenie „Oj!”, i x jest w stanie wewnętrznym P).⁸

W wariancie stwierdzającym **identyczność bólu i pewnej własności funkcjonalnej**: doznawanie bólu = bycie w pierwszym z dwóch stanów wewnętrznych, z których pierwszy jest przyczynowo spowodowany przez ukłucie szpilką, a jednocześnie przyczynowo powoduje drugi stan wewnętrzny i wypowiedzenie „Oj!”.

Przykład maszyny do określania parzystości pozwala zilustrować główną ideę funkcjonalizmu w kwestii tego, czym są stany mentalne.⁹



Maszyna „stwierdza”, czy pokazano jej „na wejściu” parzystą, czy też nieparzystą liczbę jedynek (traktuje ona „0” jako parzyste). Posiada ona możliwość identyfikacji

⁶ J. Fodor, *Special Sciences (or: Disunity of Science as a Working Hypothesis)*, „Synthese”, v. 28 (1974), s. 97-115.

⁷ N. Block, *op. cit.*, oraz: tenże, *Troubles with Functionalism* (1978), przedruk [w:] A. Goldman, (red.), *Readings in Philosophy and Cognitive Science*, Cambridge (Massachusetts) 1993, The MIT Press, s. 231-253.

⁸ N. Block, *Functionalism*, s. 190.

⁹ *Ibid.*, s. 189.

dwóch stanów wejściowych „0” i „1”, ma dwa możliwe stany wewnętrzne S_1 i S_2 oraz posiada możliwość wykazywania dwóch stanów wyjściowych „Parzyste” i „Nieparzyste”. Każdy z czterech prostokątów określa, jaki stan należy wykazać „na wyjściu” i do jakiego stanu wewnętrznego należy przejść w kolejnym kroku, jeśli dane są aktualny stan wejściowy i aktualny stan wewnętrzny. Na przykład, można przyjąć, że górny lewy prostokąt określa — za pomocą następujących dwóch zdań warunkowych — wartości dwóch odpowiednich funkcji:

(1) jeśli maszyna jest w wewnętrznym stanie S_1 i identyfikuje „1” na wejściu, to wykazuje ona „**Nieparzyste**” na wyjściu;

(2) jeśli maszyna jest w wewnętrznym stanie S_1 i identyfikuje „1” na wejściu, to przechodzi ona do wewnętrznego stanu S_2 .

Postawmy następujące pytanie: czym właściwie jest stan wewnętrzny, np. S_1 , w przypadku maszyny do określania parzystości? Można na nie odpowiedzieć, że status stanu S_1 w pełni polega na jego **relacjonalnym charakterze** i jest całkowicie zadany przez powyższy diagram (lub też rodzaj macierzy). Opis stanu wewnętrznego S_1 stanowi osiem branych łącznie zdań warunkowych, wyrażających m.in. to, że będąc w stanie S_1 i identyfikując „1” jako stan wejściowy, maszyna przechodzi do innego stanu wewnętrznego, który można charakteryzować w podobny sposób.

Charakterystyka S_1 może również zostać podana *explicite* przez następującą definicję sformułowaną w języku rachunku kwantyfikatorów II rzędu:

x jest w $S_1 \Leftrightarrow \exists P \exists Q$ [(jeśli x jest w P i identyfikuje „1” na wejściu, to x przechodzi do Q i wykazuje „**Nieparzyste**” na wyjściu) oraz (jeśli x jest w Q i identyfikuje „1” na wejściu, to x przechodzi do P i wykazuje „**Parzyste**” na wyjściu) oraz (...) oraz x jest w P].

W powyższej definicji pominięliśmy dla uproszczenia analogiczne warunki zobrażone przez dolną połowę diagramu.

Sformułować można następujące tezy dotyczące funkcjonalizmu i stanów funkcjonalnych:

(1) Według funkcjonalizmu, status wewnętrznych stanów umysłowych jest taki sam, jak status stanów wewnętrznych opisanej wyżej maszyny do określania parzystości. Status ten wyczerpuje się całkowicie przez podanie związków danego wewnętrznego stanu umysłowego z innymi stanami wewnętrznymi, ze stanami wejściowymi i wyjściowymi;

(2) Pod tym „obliczeniowym” względem umysł jest taki jak scharakteryzowany wyżej automat (maszyna Turinga), a jego stany wewnętrzne (stany mentalne, umysłowe) definiować można za pomocą podobnych definicji.

(3) S_1 i S_2 są **stanami (własnościami) drugiego rzędu**, ponieważ polegają na posiadaniu innych stanów (własności), np. mechanicznych lub elektronicznych, wchodzących w odpowiednie relacje między sobą. Te inne stany (własności), po których przebiega kwantyfikacja w odpowiednich definicjach, określane są jako **realizacje stanów (własności) funkcjonalnych**.

(4) Jeden i ten sam stan funkcjonalny (np. S_2) może zostać zrealizowany na różne sposoby, przez stany mechaniczne, elektroniczne itd. Ponieważ ten sam stan funkcjonalny (np. S_1) może być zrealizowany w bardzo różny sposób, nie może być on utożsamiany ani ze stanem mechanicznym, ani ze stanem elektronicznym itd. Funkcjonalizm wyklucza zatem skrajny fizykalizm (teorię identyczności).

(5) Taki sam stan fizyczny może być realizatorem różnych stanów funkcjonalnych w różnych maszynach Turinga.¹⁰

Ogólna postać zrelatywizowanej do pewnej teorii T definicji stanu psychicznego (np. bólu) jest następująca.¹¹ Przyjmijmy, że T jest teorią psychologiczną posługującą się n terminami oznaczającymi stany psychiczne, z których np. dwunasty to stan bólu. Wtedy

$$x \text{ doznaje bólu} \Leftrightarrow \exists F_1, \dots, \exists F_n [T(F_1, \dots, F_n) \ \& \ x \text{ jest w stanie } F_{12}].$$

W ramach funkcjonalizmu charakteryzujemy więc stany psychiczne w terminach nie-mentalnych. Posługujemy się przy tym kwantyfikacją po zmiennych odnoszących się do **realizacji** stanów (własności) psychicznych, a nie do nich samych. Zakładamy wewnętrzną, teoretyczną strukturę — siatkę związków przyczynowych — połączoną z rzeczywistością empiryczną tylko poprzez powiązania ze stanami wejściowymi i wyjściowymi. Można odwołać się tu do pojęcia **zdania Ramsey’owskiego teorii T** i nazwać wyrażenie występujące po prawej stronie powyższej definicji **Ramsey’owskim funkcjonalnym korelatem stanu bólu względem teorii T**.

Wyróżnić można dwie odmiany funkcjonalizmu. Pierwsza z nich to **psychofunkcjonalizm**¹² (gdy T jest empiryczną, naukową teorią psychologiczną) zapoczątkowany w latach 60. m.in. przez H. Putnama i J. Fodora, którzy kierowali funkcjonalistyczną argumentacją przeciwko skrajnemu fizykalizmowi (teorii identyczności), ujmując stany psychiczne jako stany wewnętrzne jakiegoś urządzenia obliczeniowego (np. tzw. maszyny Boltzmann). Inna odmiana określona zostaje jako **funkcjonalizm pojęciowy**¹³ (konceptualny), zainicjowany przez D. Armstronga i D. Lewisa, którzy zaproponowali funkcjonalistyczną analizę znaczenia terminów mentalnych. Według tego ujęcia, nie rozumie znaczenia terminu ‘ból’ ten, kto nie wie, że w zwykłych warunkach działamy tak, żeby unikać bólu.

Block uważa, iż teorie umysłu mogą dawać odpowiedź na pytanie „ontologiczne” (co istnieje?) lub na pytanie „metafizyczne” (czym są przedmioty danego typu, a w szczególności, co mają wspólnego ze sobą różne bóle, na mocy czego są one bólami?)¹⁴.

Funkcjonalista odpowiada na pytanie „metafizyczne”, nie dając odpowiedzi na pytanie „ontologiczne”. Wspólna dla różnych rodzajów bólu jest ich funkcjonalna

¹⁰ *Ibid.*, s. 189-190.

¹¹ *Ibid.*, s. 190.

¹² N. Block, *Troubles with ...*, s. 235.

¹³ N. Block, *Functionalism*, s. 190.

¹⁴ *Ibid.*, s. 191.

(przyczynowa) struktura, natomiast nie rozstrzyga się tego, czy byty doznające bólu mają jakieś nie-fizyczne składniki. Aby być takim automatem, jak opisana wyżej maszyna do określania parzystości, rzeczywiste konkretne urządzenie musi jedynie posiadać takie same zewnętrzne i wewnętrzne stany, powiązane przyczynowo ze sobą w przedstawiony sposób. Opis takiej maszyny nie mówi nam nic na temat tego, z czego jest ona zbudowana lub jak działa; nie wyklucza on np., że za jej funkcjonalną organizację mogłaby być odpowiedzialna niematerialna dusza.

Z punktu widzenia funkcjonalizmu fizykalizm występuje w dwóch wersjach, skrajnej i umiarkowanej.¹⁵ Pierwsza z nich — **fizykalizm typologiczny** — głosi „metafizyczną” tezę o fizykalnej naturze **typów** stanów psychicznych i ma być całkowicie niezgodna z funkcjonalizmem. Druga — **fizykalizm egzemplaryczny** — ogranicza się do tezy „ontologicznych”, które można uzgodnić z funkcjonalizmem, np., że każdy rzeczywisty, konkretny przedmiot składa się w całości z fizycznych cząstek elementarnych. Możliwy jest zatem umiarkowanie fizykalistyczny funkcjonalizm, dopuszczający tylko fizyczne realizacje, tzn. spełniający wymóg, żeby wszystkie własności lub stany, po jakich kwantyfikujemy w definicjach stanów psychicznych, były fizyczne.

Obecnie chciałbym przypomnieć¹⁶ ogólną linię argumentacji Fodora, przedstawioną w artykule *Special Sciences (or Disunity of Science as a Working Hypothesis)*¹⁷, która prowadzi do wniosków przeciwstawiających się postulatowi redukcjonistycznej jedności nauki (tym samym skierowana jest również przeciwko skrajnemu fizykalizmowi, czyli teorii identyczności). W ramach tej argumentacji sformułowana zostaje teza, iż **predykaty przypisane rodzajom naturalnym występujące w naukach szczegółowych nie odpowiadają predykatom przypisanym rodzajom naturalnym z zakresu podstawowej nauki redukującej (fizyki), w szczególności** — nie są z nimi koekstensywne; a klasyfikacje oparte na tych predykatkach w zasadniczo nieusuwalny sposób krzyżują się ze sobą. Teza ta jest uzasadniana za pomocą przykładów z różnych dyscyplin naukowych (np. ekonomii lub ekologii) oraz przez odwołanie się do funkcjonalistycznego ujęcia stanów umysłu (stanów psychicznych) w ramach tzw. obliczeniowej teorii umysłu, według której umysł człowieka lub zwierzęcia może być opisany, na przykład, jako pewien rodzaj automatu probabilistycznego, czyli uogólnionej maszyny Turinga (maszyny Boltzmann).

Neopozytywistyczny program jedności nauki przedstawiony został w 1929 r. w sposób ogólny przez Carnapa, Hahna i Neuratha w manifestie Koła Wiedeńskiego (*Wissenschaftliche Weltauffassung: Der Wiener Kreis*), a następnie był stopniowo uszczegółowiany. Na przykład w 1938 r. Carnap wymieniał dwa postulaty — „jedności języka” i „jedności praw”.¹⁸

¹⁵ W. Strawiański, *op. cit.*, s. 196.

¹⁶ *Ibid.*, s. 195.

¹⁷ J. Fodor, *op. cit.*

¹⁸ R. Carnap, *Logical Foundations of the Unity of Science*, [w:] O. Neurath, R. Carnap, C. Morris, (red.), *Encyclopedia and Unified Science. International Encyclopedia of Unified Science, v. 1, no 1*, Chicago 1938, The University of Chicago Press, s. 42-62. Por. też: W. Strawiański, *op. cit.*, s. 28-29.

— **Jedność języka:** miała to być definicyjna redukcja terminologii nauk szczegółowych do ugruntowanego empirycznie słownika fizykalistycznego lub fenomenalistycznego;

— **Jedność praw:** miało to być wyprowadzenie praw psychologii i nauk społecznych z praw fizyki i biologii; „konstrukcja jednego systemu praw dla całej nauki jako cel przyszłego jej rozwoju”.

Osiągnięcie takiego stadium „nauki zjednoczonej” (*Einheitswissenschaft, Unified Science*) miałyby duże znaczenie dla zagadnienia psychofizycznego, ponieważ zniesienie odrębności przyrodoznawstwa i nauk społeczno-humanistycznych pozwoliłoby przewyciężyć przeciwstawianie sobie tego, co fizyczne i tego, co psychiczne. Z kolei program integracji nauk szczegółowych na poziomie całych teorii był zaproponowany przez Feigla¹⁹ i został przedstawiony bardziej szczegółowo w artykule napisanym przez Oppenheima i Putnama.²⁰ Ten ostatni artykuł hipotetycznie postulował redukcjonistyczny obraz zarówno wiedzy naukowej, jak i wszechświata.

Fodor zaproponował zrewidowanie redukcyjnej wersji postulatu jedności praw nauki właściwej logicznemu empiryzmowi.²¹ Według niego, redukcjonizm jest niewłaściwą interpretacją ogólności fizyki jako nauki podstawowej. Fodor wprowadza wspomniane rozróżnienie między jednostkowym, **egzemplarycznym fizykalizmem** i **fizykalizmem typologicznym**, rodzajowym. Redukcjonizm zakłada według niego fizykalizm egzemplaryczny (*token physicalism*), zgodnie z którym wszystkie zdarzenia opisywane przez naukę to zdarzenia fizyczne; ponadto zakłada również odpowiedniość (np. koekstensywność) predykatów wyrażających rodzaje naturalne różnych nauk i fizyki, czyli fizykalizm typologiczny (*type physicalism*). Predykaty wyrażające rodzaje naturalne danej nauki wyróżnia to, że odnoszą się one do zakresów zmiennych związanych we właściwych prawach tej nauki. Dzięki założeniu odpowiedniości predykatów przypisywanych rodzajom naturalnym różnych dziedzin, redukcjonizm może postulować istnienie praw pomostowych między przedmiotami tych dziedzin.

Fodor to odrzuca, kwestionując możliwość realizacji postulatu jedności praw. Według niego, predykaty wyrażające rodzaje naturalne w taksonomiach nauk szczegółowych i nauki podstawowej nie są koekstensywne; mogą między nimi co najwyżej zachodzić związki dysjunkcyjne, niedające podstawy do formułowania właściwych praw pomostowych. Postulaty pomostowe w indywidualnych schematach redukcyjnych wyrażają, według Fodora, kontyngentną, przypadkową identyczność zdarzeń i nie mają charakteru praw. Ilustracją dysjunkcyjnego charakteru wspomnianych

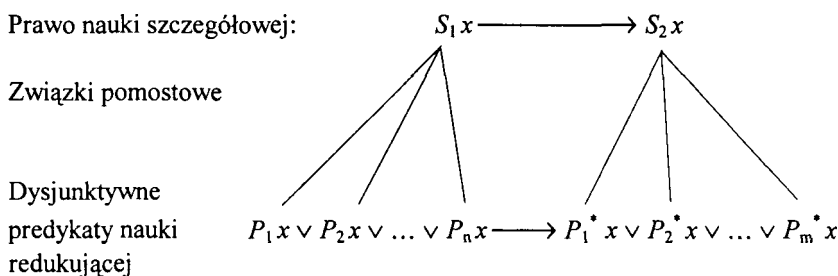
¹⁹ H. Feigl, *Some Major Issues and Developments in the Philosophy of Science of Logical Empiricism*, [w:] H. Feigl, M. Scriven, (red.), *The Foundations of Science and Concepts of Psychology and Psychoanalysis*, MSPS v. 1, Minneapolis 1956, University of Minnesota Press, s. 3-37.

²⁰ P. Oppenheim, H. Putnam, *Unity of Science as a Working Hypothesis*, [w:] H. Feigl, M. Scriven, G. Maxwell, (red.), *Concepts, Theories and the Mind-Body Problem*, MSPS vol. 2, Minneapolis 1958, University of Minnesota Press, s. 3-36.

²¹ J. Fodor, *op. cit.*, s. 97.

związków może być predykat „wymiana monetarna”, występujący w ekonomii np. w tzw. prawie Greshama. Każda wymiana monetarna jest jakimś zdarzeniem fizycznym, ale trudno przypuścić, by zdarzenia takie łączyła jakaś istotna prawidłowość fizyczna; dla ich ogólnej, fizycznej charakterystyki pozostaje jedynie dysjunktywne „zsumowanie” fizycznych charakterystyk wymian różnego typu.

Poniżej umieściliśmy diagram przedstawiający według Fodora związki między nauką redukowaną i redukującą²²:



Prawa nauki redukującej: $P_1x \rightarrow P_2^*x, P_2x \rightarrow P_m^*x, \dots, P_nx \rightarrow P_1^*x$;

Argumentacja Fodora niewątpliwie znajdowała oparcie w funkcyjalistycznym ujęciu stanów psychicznych przez Putnama.²³ Stany psychiczne takie jak przeżywanie bólu, złości, pragnienie czegoś, a także bycie o czymś przekonany, charakteryzują się w tym ujęciu **posiadaniem pewnej organizacji funkcjonalnej**, polegającej na powiązaniu ich z przetwornikami sensorycznymi („wejściami”) i motorycznymi („wyjściami”), a także na związku z innymi stanami psychicznymi. Z tej perspektywy stan taki jak ból nie jest identyfikowany z fizyczno-chemicznym stanem układu nerwowego, lecz jest stanem innego rodzaju — stanem funkcjonalnym całego organizmu.

W ramach proponowanej przez Putnama koncepcji funkcyjalistycznej prawa psychologiczne miałyby być wyprowadzane z funkcjonalnych opisów zachowania się istniejących organizmów oraz założenia o identyfikacji stanów psychicznych i funkcjonalnych. Organizmy mogą być ponadto „funkcyjalnie izomorficzne” niezależnie od ich struktury fizyczno-chemicznej, co stanowić może podstawę do wyróżnienia wśród właściwych im stanów odpowiednich rodzajów naturalnych, opisywanych przez prawa funkcjonalne (np. wspólnego stanu bólu u człowieka i u ośmiornicy). Co więcej, teza o funkcjonalnym izomorfizmie może zostać rozszerzona tak, że obejmować będzie także automaty, komputery i dowolne inne systemy o właściwej organizacji funkcjonalnej. W ramach takiej „obliczeniowej koncepcji umysłu” przyjmuje się, że **dla każdego organizmu**, któremu można przypisywać stany psychiczne, **istnieje jego najlepszy funkcjonalny opis jako automatu probabilistycznego**, a stany psy-

²² *Ibid.*, s. 109.

²³ H. Putnam, *The Mental Life..., The Nature of Mental...*

chiczne tego organizmu mogą być utożsamione ze stanami urządzenia sterującego tego automatu (zrelatywizowanymi do opisu).

Czy wobec powyższych rozważań możliwa jest redukcja psychologii do neurologii? Możliwość zredukowania psychologii do neurologii zachodziłaby, gdyby każdy predykat wyrażający jakiś rodzaj naturalny psychologii znajdował swój odpowiednik na poziomie neurologicznym, a związek ten miałby charakter prawa pomostowego. Według Putnama wiele wskazuje jednak na to, że *układy nerwowe wyższych organizmów mogą osiągać dany stan psychiczny na bardzo dużo różnych sposobów*. W świetle tego jest wątpliwe, czy istnieją neurologiczne rodzaje naturalne koekstensywne z psychologicznymi rodzajami naturalnymi. W rezultacie, *nawet jeśli każdy konkretny przypadek zgodnego z pewnym prawem warunkowania stanu psychicznego jakiegoś typu przez stan psychiczny innego typu będzie miał u podłoża określone następstwo dwóch konkretnych zdarzeń neurologicznych, to nie będzie ono instancją dostatecznie ogólnego prawa neurologicznego*. W grę wchodzić będzie co najwyżej prawdziwa empiryczna generalizacja, której poprzednik i następnik będą miały postać niejednorodnej dysjunkcji. Ma to prowadzić do odrzucenia postulatu redukcjonistycznej jedności odpowiednich praw i teorii, gdyż predykaty wyrażające rodzaje naturalne w taksonomiach psychologii i neurologii nie będą koekstensywne.

W ten sposób scharakteryzowana wyżej, funkcjonalistyczna, obliczeniowa teoria umysłu miała popierać argumentację Fodora dotyczącą odrębności i autonomii poszczególnych dyscyplin naukowych. Odnotujmy jeszcze, że mimo wszystko Fodor utrzymuje pewną zrewidowaną wersję postulatu jedności nauki, pisząc: „[...] jest wystarczające dla celów jedności nauki, by każde prawo nauk szczegółowych było redukowalne do fizyki za pomocą zdań pomostowych wyrażających prawdziwe empiryczne generalizacje”²⁴

²⁴ J. Fodor, *op. cit.*, s. 108.