

Stanisław Judycki

**Współczesne odmiany materializmu:  
identyczność psychofizyczna, superweniencja,  
eliminacjonizm\***

Za źródła XX-wiecznej teorii identyczności psychofizycznej uznaje się program filozoficzny Koła Wiedeńskiego (szczególnie jego tezę weryfikacjonistyczną<sup>1</sup>), rozwój badań neurofizjologicznych, filozoficzną interpretację psychologii behawiorystycznej<sup>2</sup> oraz, w ostatnich czasach, pojawienie się komputerów i tzw. nauk komputerowych. Zwykle wymienia się w tym kontekście M. Schlicka, a za początkodawców identyzmu w

---

\* Tekst poniższy został opublikowany w: „Zeszyty Naukowe KUL”, 38 (1995), nr 3-4, s. 49-61.

<sup>1</sup> Weryfikacjonizm był teorią głoszoną pierwotnie przez przedstawicieli tzw. logicznego empiryzmu. Semantyka weryfikacjonistyczna nakazywała identyfikowanie znaczenia zdań należących do nauki ze znaczeniem zdań obserwacyjnych. Weryfikacja empiryczna w powyższym sensie miała udzielać zdaniom nie tylko waloru naukowego obowiązywania, lecz także nadawać im znaczenie. Zgodnie więc z tą zasadą, jeżeli stany mentalne są ludziom przypisywane w oparciu o ich cielesne zachowania, a więc jeżeli zdania dotyczące tych stanów są weryfikowane tylko w oparciu o dające się zmysłowo ująć oznaki fizyczne, to pojęcia mentalne w rzeczywistości oznaczają stany cielesne. Według tego stanowiska, znaczenie zdania jest całkowicie określone przez jego konsekwencje empiryczne, a stąd też zdanie to jest całkowicie przekładalne na zdania o tych konsekwencjach. Tak radykalne podejście wywołało dyskusję na temat, co należy uznać za odpowiednią metodę weryfikacji zmysłowej w przypadku nauk przyrodniczych oraz na temat konsekwencji wynikających z próby odniesienia tej zasady do samej siebie. Frustracja niemożliwością rozwiązania tych problemów doprowadziła C. Hempela do agnostycyzmu co do możliwości odkrycia odpowiedniej dla nauki zasady weryfikacjonistycznej; por. T. Burge, *Philosophy of Language and Mind: 1950-1990*, „Philosophical Review” 101 (1992), s. 5.

<sup>2</sup> Behawioryzm psychologiczny (naukowy) jest tezą, że wyłącznym przedmiotem zainteresowania psychologii ma być przewidywanie zachowania (B. F. Skinner). Ten rodzaj behawioryzmu traktował odwoływanie się do subiektywnych stanów mentalnych za bezużyteczne dla psychologii jako nauki, natomiast behawioryzm filozoficzny (logiczny) uważał samą ideę istotnie prywatnych stanów mentalnych za ideę logicznie absurdalną. Głównym propagatorem tego stanowiska był G. Ryle w książce *The Concept of Mind* (1949). Przedmiotem krytyki Ryle'a stał się dualizm kartezjański, przyjmujący - wedle jego interpretacji - istnienie danych w introspekcji, istotnie prywatnych, niekorygowalnych i osadzonych w niematerialnej substancji stanów mentalnych. Według Ryle'a terminy mentalne nie odnoszą się do tego rodzaju stanów, lecz do dającego się obserwować zachowania się jakiegoś podmiotu. I tak np., gdy mówimy o kimś, że jest próżny, to nie sądzi my, że znajduje się w nim pewien osobliwy stan próżności, zakotwiczony w duchowym materiale, lecz uznajemy, że w jego zachowaniu da się wykryć pewien stały wzorec.

latach 50. naszego wieku uchodzą U.T. Place, J.J. Smart, H. Feigl<sup>3</sup>. Generalne motywy prowadzące do stanowiska materializmu identycznościowego dadzą się podzielić w następujący sposób: 1. motywy wypływające z ogólnych założeń fizykalistycznych, 2. motywy biorące się z pewnych analiz dotyczących sposobu funkcjonowania naszego aparatu pojęciowego w dziedzinach poza psychologią i neurofizjologią i z przeniesienia wyników tych analiz do filozofii umysłu, 3. motywy opierające się na analogiach między udanymi redukcjami interteoretycznymi poza psychologią a możliwą (przyszłą) redukcją tego, co psychiczne do tego, co fizyczne, 4. racje empiryczne, odwołujące się do wyników badań neurofizjologicznych (i nauk im pokrewnych) wraz z projekcjami materialistycznych konsekwencji, które dla filozofii umysłu będzie miał rozwój tych nauk w przyszłości. U autorów piszących na temat filozofii umysłu wszystkie te motywy prowadzące do identyzmu są faktycznie ze sobą splecione i często trudno jest je od siebie wyraźnie oddzielić.

Jeżeli chodzi o pierwszy typ argumentacji, tzn. na podstawie ogólnych założeń fizykalistycznych, to należy odróżnić między fizykalizmem jako postulatem metodologicznym a fizykalizmem filozoficznym (esencjalnym). Tezę fizykalizmu filozoficznego można sformułować następująco: wszystkie obiekty albo są identyczne z obiektami fizycznymi, albo są z nich złożone; inaczej mówiąc, dla każdej zrealizowanej w świecie własności istnieją warunki fizyczne wyjaśniające jej realizację oraz dla niej wystarczające; wszystkie bazowe prawa są prawami fizyki. Fizykalizm metodologiczny polega tylko na postulacie, aby dla wszystkich fenomenów poszukiwać wyjaśnień fizycznych. Dopiero gdy fizykalizm zinterpretujemy jako tezę esencjalną (filozoficzną), dyskusja na temat relacji między umysłem a ciałem będzie jedną z dróg weryfikujących prawdziwość tego stanowiska. Jeżeli w ramach tej dyskusji ocena fizykalizmu wypadnie negatywnie, wtedy oczywiście pozostaną inne dziedziny weryfikujące tezę fizykalistyczną: biologia, ekonomia, obiekty abstrakcyjne.

Dla fizykalizmu filozoficznego można szukać poparcia w sukcesach pragmatycznych, do jakich doprowadziły wyjaśnienia czysto fizyczne. Pytaniem jest jednak, czy to, co przynosi sukces w jednej dziedzinie, tzn. w dziedzinie wyjaśniania fenomenów świata pozaumysłowego i pozaludzkiego, musi również doprowadzić do sukcesu eksplanacyjnego w innych dziedzinach. Czy więc lepsze efekty odnosić będzie w przyszłości psychologia zorientowana mentalistycznie, czy też raczej jakaś przyszła psychologia materialistyczna. Współczesna psychologia i. wszystkie oparte na niej terapie jest psychologią mentalistyczną, tzn. posługuje się dualistycznym aparatem pojęciowym przyjmując (obok oddziaływań czysto fizycznych) istnienie fizycznie nieredukowalnych oddziaływań między

<sup>3</sup> Chodzi tu o następujące teksty: M. Schlick, *Allgemeine Erkenntnislehre*, Berlin 1925<sup>2</sup>; U.T. Place, *Is Consciousness a Brain Process?*, „British Journal of Psychology” 47 (1956), s. 44-50; J.J. Smart, *Sensations and Brain Processes*, „Philosophical Review” 68 (1959), s. 141-156; H. Feigl, *Principles and Problems of Theory Construction in Psychology*, w: D. Dennis (wyd.), *Current Trends in Psychological Theory*, Pittsburgh 1951, s. 179-213; H. Feigl, *The 'Mental' and the 'Physical'*, w: H. Feigl, M. Scriven, G. Maxwell (wyd.), *Concepts, Theories and the Mind-Body Problem* (Minnesota Studies in the Philosophy of Science. T. II). Minneapolis 1958, s. 370-497.

własnościami mentalnymi.

Najogólniejszą, czysto pojęciową, analizę, która prowadzi do teorii identyczności psychofizycznej, można streścić w następujący sposób. Pojęcia neurofizjologiczne (odnoszące się do zdarzeń w systemie nerwowym i w mózgu człowieka) i pojęcia mentalistyczne (utworzone w oparciu o introspekcję) posiadają różną treść (intensję), lecz są ekstensjonalnie jednakowe. Zdania na temat zdarzeń psychicznych i zdania dotyczące zdarzeń neurofizjologicznych odnoszą się więc faktycznie do tych samych przedmiotów. Człowieka można opisać bądź jako zwierzę rozumne, bądź jako nieopierzonego dwunoga - te dwa opisy mają wprawdzie różne znaczenia, lecz odnoszą się do tej samej klasy istot żywych. Tego rodzaju identyczność nie ma charakteru analitycznego i koniecznego (tak jak analityczna identyczność między np. pojęciami 'kawaler' – 'mężczyzna nieżonaty'), lecz jest to identyczność kontyngentna, oparta o doświadczenie, jak np. między 'woda' a 'H<sub>2</sub>O', czy między 'temperatura' a 'średnia energia kinetyczna'. Według wersji teorii identyczności nazywanej identycznością typiczną (*Type-Type-Identity*) rodzaje mentalne (np. przekonania, życzenia) należy zidentyfikować z określonymi typami zdarzeń neurofizjologicznych w mózgu. Np. radość byłaby ekstensjonalnie równoważna z elektrycznym pobudzeniem określonego typu włókien nerwowych *N*.

Oslabioną wersją teorii identyczności jest teoria identyczności egzemplarycznej (*Token-Identity-Thesis*), w której identyfikuje się szczegółowe pojęcia psychologiczne (lub szczegółowe zdarzenia psychiczne) ze szczegółowymi pojęciami neurofizjologicznymi (lub ze szczegółowymi zdarzeniami neuronalnymi). U osoby *A* radość byłaby więc identyczna z pobudzeniem włókien nerwowych *N* w czasie *t*<sub>1</sub>, podczas gdy u osoby *B* (lub u osoby *A* w czasie *t*<sub>2</sub>) korelat neurofizjologiczny radości mógłby być całkiem inny. W identyzmie egzemplarycznym chodzi nie o generalną, lecz o partykularną identyczność, nie o identyczność interpersonalną, lecz o identyczność intrapersonalną. Jednemu typowi zdarzeń psychicznych nie musi więc odpowiadać dokładnie jeden typ zdarzeń neurofizjologicznych. Jeden i ten sam rodzaj zdarzeń psychicznych może mieć różne realizacje neurofizjologiczne.

Identyfikacja zdarzeń i procesów mentalnych ze zdarzeniami i procesami neurofizjologicznymi motywowana jest przykładami udanych redukcji interteoretycznych w innych dziedzinach wiedzy, których efektem ostatecznym ma być redukcja ontologiczna. Najczęściej podawanym przykładem jest redukcja tzw. termodynamiki fenomenologicznej do mechaniki statystycznej. W jej wyniku pojęciom 'ciśnienie' i 'temperatura' zostało przyporządkowane (w gazie idealnym) wyrażenie 'średnia energia kinetyczna molekuł'. Przyporządkowanie to nie polega jednak na połączeniu jednego (každorazowego w danej chwili) mikrostanu molekularnego z jednym (každorazowym) stanem termicznym. Urzeczywistnienie określonego przez konkretną temperaturę makrostanu da się osiągnąć przez różne kombinacje poszczególnych energii kinetycznych. Istnieje więc mnogość mikrostanów, które z punktu widzenia 'makro' są sobie równoważne. Można zatem mówić o funkcjonalnej równoważności mikrostanów z danym makrostanem.

Samo zagadnienie charakteru redukcji interteoretycznej nastęca wiele trudności i stanowi osobny problem w filozofii nauki<sup>4</sup>. W dyskusji nad zagadnieniem umysł - ciało ważna jest jednak sama idea, opierająca się na redukcji interteoretycznej, redukcji ontologicznej, tzn. zastąpienia jednego rodzaju bytów innym rodzajem (czy też jednego rodzaju pojęć innymi pojęciami). Oczywiście, kierunek tych redukcji jest generalnie Ockhamowski, a dokładniej biorąc, fizykalistyczny. Argumentacja oparta o istnienie udanych redukcji interteoretycznych posiada (w oczach jej zwolenników) też tę zaletę, że przeciwników wynikających z niej konsekwencji ontologicznych wydaje się ustawiać w roli dogmatycznych przeciwników postępu naukowego<sup>5</sup>.

Redukcje interteoretyczne mogą mieć charakter ‘miękki’ lub ‘twardy’<sup>6</sup>. Redukcje miękkie są redukcjami eksplanacyjnymi. Teoria  $T_1$  redukuje się międko do teorii  $T_2$ , jeżeli jej sukces eksplanacyjny i prognostyczny są wyjaśniane przez teorię  $T_2$ . Zazwyczaj redukcje tego rodzaju wykazują dominację korelacji typ-typ (czyli dominację praw) między przedmiotami referencji redukowanej i redukującej teorii. Im bardziej ‘międką’ jest redukcja, tym bardziej ścisłe będą korelacje typ-typ między redukowaną a redukującą teorią. Np. optyka fizyczna została międko zredukowana do teorii elektromagnetycznej, gdyż nie tylko jej efekt eksplanacyjny, lecz także jej założenia i prawa otrzymały głębsze i bogatsze wyjaśnienie w teorii elektromagnetycznej. Na przeciwległym krańcu znajdują się redukcje twarde, nazywane też eliminacyjnymi. Twardo redukująca teoria eliminacyjnie wyjaśnia przedmioty, które akceptowała teoria redukowana.

<sup>4</sup> Standardowy model redukcji interteoretycznej został sformułowany przez E. Nagla i składa się z dwóch warunków formalnych: 1. wszystkie nieobecne w teorii redukowanej  $T_1$  pojęcia teorii redukującej  $T_2$  muszą zostać połączone z pojęciami teorii  $T_1$  za pomocą tzw. reguł korespondencji; 2. za pomocą tych reguł wszystkie prawa  $T_1$  muszą być wyprowadzalne z praw  $T_2$ ; por. E. Nagel, *The Structure of Science. Problems in the Logic of Scientific Explanation*. London 1961 (pol. tłum. *Struktura nauki. Zagadnienia logiki wyjaśnień naukowych*. Warszawa 1970 s. 294-343). Do nieformalnych warunków redukcji należy to, że założenia teorii redukującej muszą mieć oparcie empiryczne, że przeprowadzana redukcja powinna doprowadzać do korektury znanych dotąd praw empirycznych, powinna ukazywać zaskakujące powiązania między tymi prawami i przewidywać całkiem nowe prawa. Trudność leży w tym, że aby spełnić drugi warunek formalny, warunek wyprowadzalności, teoria redukowana i teoria redukująca muszą być ze sobą zgodne, aby zaś spełnić warunek nieformalny teoria redukująca musi poprawiać teorię redukowaną, co z kolei zakłada, że obie teorie są ze sobą niezgodne. Korektura praw  $T_1$  przez  $T_2$  uniemożliwiałaby zatem spełnienie warunku wyprowadzalności. Rozwiązanie tej trudności niektórzy widzą we wprowadzeniu do teorii redukcji interteoretycznej pojęcia interpretowalności w sensie A. Tarskiego; por. M. Carrier, J. Mittelstraß, *Geist, Gehirn, Verhalten. Das Leib-Seele-Problem und die Philosophie der Psychologie*, Berlin-New York 1989 s. 47-48.

<sup>5</sup> „The idea is to make you, the sceptical reader, feel that if you don't believe the view being advanced, you are playing Cardinal Bellarmine to the author's Galileo. Other favorites are phlogiston and vital spirits, and again the idea is to bully the reader into supposing that if he or she doubts, for example, that computers are actually thinking, it can only be because the reader believes in something as unscientific as phlogiston or vital spirits” (J.R. Searle, *The Rediscovery of the Mind*, Cambridge MA, 1992, s. 5).

<sup>6</sup> Por. K.V. Wiles, *Mind and Body: Some Forms of Reductionism*, w: G.H.R. Parkinson (wyd.), *An Encyclopaedia of Philosophy*, London 1988, s. 402-426.

Najczęściej podawanym tu przykładem jest redukcja flogistonowej teorii spalania do teorii tlenowej. W wyniku tej redukcji flogiston nie został odniesiony za pomocą korelacji typ-typ do czegoś z zakresu teorii tlenowej. To, co miał wyjaśniać flogiston, zostało w lepszy i odmienny sposób wyjaśnione przez teorię tlenową. Nowa teoria wyeliminowała z nauki pojęcie flogistonu. Nie istnieje jednak żadna konieczność, aby teoria redukująca twardo (w tym wypadku teoria tlenowa), wyjaśniała teorię przez siebie zastępowaną, tzn. aby wyjaśniała przyjmowane przez starą teorię przedmioty, byty czy procesy i ich sposób zachowania się. Analogicznie do tego mielibyśmy dwie wersje fizykalizmu: albo przyszła teoria neurofizjologiczna wyjaśni miękko potoczną psychologię introspekcyjną, tzn. zachowując jej terminy i prawa, ale jednocześnie dostarczy terminów i praw, które wyjaśniają więcej fenomenów mentalnych i zrobią to głębiej niż dotychczasowa teoria; albo przyszła redukcja będzie miała charakter eliminacyjny, tzn. pojęcia i prawa psychologii potocznej (mentalistycznej) zostaną zastąpione przez pojęcia i prawa neurofizjologii. W wyniku tej redukcji mówienie o fenomenach mentalnych będzie miało tyle sensu, co mówienie o czarownicach zamiast o neurotycznych kobietach.

Oprócz rozważań nad znaczeniem i referencją terminów psychologicznych i neurofizjologicznych oraz ogólnych porównań z udanymi redukcjami interteoretycznymi, argumenty za identycznością psychofizyczną czerpią swoją motywację z racji empirycznych bądź quasi-empirycznych. Racjami empirycznymi miałyby być dotychczas stwierdzone korelacje między zdarzeniami neuronalnymi (cielesnymi) a zdarzeniami świadomymi. Korelacje te są zauważane już na poziomie doświadczenia potocznego, np. w postaci wpływu, który na stany świadome wywierają różne środki halucynogenne. Jednak właściwym przedłużeniem tych prostych obserwacji są systematyczne badania anatomiczne stwierdzające np. odpowiedniości między pobudzeniami różnych obszarów mózgu a różnymi reakcjami i funkcjami świadomymi. Quasi-empiryczną ekstrapolację aktualnego stanu badań neurofizjologicznych na ten temat jest idea tzw. autocerebrokopu, która pochodzi prawdopodobnie od H. Feigla<sup>7</sup>. Miałby to być przyrząd, dostępny przyszłej neurofizjologii, za pomocą którego dana osoba mogłaby obserwować ścisłe korelacje między stanami własnego mózgu a własnym świadomym doświadczeniem. W związku ze stanem rzeczy prezentowanym przez autocerebrokop należałoby sięgnąć do brzytwy Ockhama i zrezygnować z tezy o istnieniu autonomicznej przyczynowości psychicznej. Stany umysłowe mogłyby być wtedy uznane za (co najwyżej) epifenomeny w stosunku do stanów fizycznych mózgu.

Teoria identyczności psychofizycznej pociąga za sobą ogromne trudności, wynikające z samego pojęcia identyczności. Sformułowana przez Leibniza tzw. zasada identyczności stwierdza, że dwie wielkości (byty, obiekty, przedmioty) są identyczne, jeżeli przysługują im takie same własności. Z tego wynika, że gdy brakuje jednakowości własności, wtedy wchodzące w rachubę przedmioty nie są identyczne. W wypadku relacji (świadomy) umysł - ciało warunek identyczności nie może być jednak spełniony. Własności mentalne, jako własności dane w introspekcji, są inne niż odpowiadające im własności fizyczne mózgu. Stojąc konsekwentnie na stanowisku identycznościowym, trzeba byłoby przypisać

<sup>7</sup> Por. H. Feigl, *The 'Mental' and the 'Physical'*, s. 456, 461.

procesom czy zdarzeniom neuronalnym np. zniecierpliwienie. Byłoby to nie tylko fałszywe, lecz również bezsensowne. Równałoby się to takiemu samemu błędowi, co próba przestrzennej lokalizacji obiektów abstrakcyjnych (np. matematycznych). W ten sposób teoria identyczności popełniłaby swoisty błąd kategorialny.

Na ten zarzut zwolennicy teorii identyczności mogą jednak odpowiedzieć, że chodzi tu tylko o szczególność naszego sposobu mówienia o zdarzeniach umysłowych, który jako taki wcale nie musi mieć żadnej naukowej doniosłości. W końcu przecież na tej samej zasadzie (przez ludzi należących do minionych pokoleń) mówienie np., że niebieskie światło błyskawicy jest tylko określoną długością fali elektromagnetycznej, musiałoby być uznane za błąd kategorialny. Jeżeli zbyt daleko posunie się zarzut błędu kategorialnego w tym sensie, to jego ofiarą padną wszystkie udane redukcje interteoretyczne w naukach. Uogólniając, można powiedzieć, że ostatecznie byłby to zarzut przeciwko zasadzie identyczności Leibniza. Zasadę tę trzeba więc ograniczyć do kontekstów ekstensjonalnych, gdyż stosowana do kontekstów intensjonalnych prowadzi do wniosku, że pojęcia o różnej intensji nigdy nie mogą mieć tej samej ekstensji.

Przejście od pierwotnej (reprezentowanej m.in. przez Smarta i Feigla) teorii identyczności typicznej do teorii identyczności egzemplarycznej spowodowane zostało zwróceniem uwagi, że jest nieprawdopodobne, aby te same procesy neurofizjologiczne (te same rodzajowo) były odpowiedzialne za pojawienie się mentalnego zdarzenia bólu zarówno w wypadku ludzi, jak i innych ssaków, ośmiornic itp.<sup>8</sup> Istnienie takiego uniwersalnie obowiązującego stanu neurofizjologicznego nie jest wprawdzie wykluczone, lecz jest mało prawdopodobne. Z podobną sytuacją będziemy mieli do czynienia, gdy weźmiemy pod uwagę istoty należące do tego samego gatunku. Ze względu na odmienność organizacji mózgu u różnych ludzi i jego przemiany w trakcie rozwoju danej jednostki, ból u osoby *A* nie musi być oparty o ten sam typ zdarzeń neuronalnych, co ból występujący u osoby *B*. Można również pytać, czy u fikcyjnych istot pozaziemskich bóle muszą być zrealizowane w oparciu o ten sam typ zdarzeń neuronalnych co u ludzi, a nawet, czy w ogóle podłoże neuronalne musi występować - inne materiały nie są przecież wykluczone.

Z tego typu rozważań wyciągnięto wniosek, że stany mentalne mają charakter logiczny (abstrakcyjny); są abstrakcyjnymi stanami funkcjonalnymi całego organizmu. Ich indywidualność jest określona przez ich połączenie z określonymi bodźcami, płynącymi ze świata, przez behawioralne reakcje na te bodźce oraz przez relacje do innych stanów umysłowych. I tak np. stan bólu może być zdefiniowany przez podanie bodźca, który go wywołał (np. okaleczenie), konsekwencje w zachowaniu, do których prowadzi (np. głośne krzyki), i relacje do innych procesów umysłowych (np. to, że powoduje pojawienie się życzenia uniknięcia bólu). Stany mentalne można więc przypisać nie tylko ludziom, lecz także całkowicie odmiennego typu organizmom, istotom pozaziemskim i komputerom. Pojęcia psychologiczne są pojęciami funkcjonalnymi - takimi, jak 'gaźnik', 'zawór' czy 'pułapka na myszy'. Rzeczy tego rodzaju identyfikuje się przez podanie funkcji, którą mają spełniać, przy czym one same mogą być zrealizowane w różny sposób. Funkcjonalistyczne podejście do umysłu nie wyklucza, że stany mentalne mogą być zrealizowane w

<sup>8</sup> Por. H. Putnam, *The Nature of Mental States* (1967), w: H. Putnam, *Mind, Language and Reality (Philosophical Papers 2)*, Cambridge 1975 s. 429-440; J. Fodor, *What Psychological States Are Not*, „Philosophical Review” 81(1972), s. 159-181.

jakimś (bliżej nieokreślonym) medium pozafizycznym (np. w czymś w rodzaju kartezyjskiej niematerialnej substancji). Funkcjonalizm nie wyklucza więc dualizmu. Funkcjonalizm stoi też na stanowisku nieeliminowalności psychologicznego aparatu pojęciowego. Warunki, które należy postawić, funkcjonalnie równoważnym, lecz anatomicznie heterogenicznym, neuronalnym realizacjom stanów psychicznych, dadzą się sformułować tylko w języku psychologicznym. Podobnie jak warunki, które należy postawić gaźnikom czy pułapkom na myszy, wymagają opisanie ich w języku funkcji, a nie w języku materiału. Pojęcia psychologiczne są niezbędne, gdyż to one muszą określać, co określony system neuronalny musi móc dokonać, aby można go było uznać za materialną realizację określonej funkcji psychicznej.

Standardowe obiekcje w stosunku do funkcjonalistycznej koncepcji umysłu dotyczą niewystarczalności opisów funkcjonalnych dla oddania jakościowego charakteru świadomości. Wymagania funkcjonalistyczne spełniałby np. całkowicie nieświadomy 'robot', który dysponowałby wszystkimi warunkami funkcjonalnymi potrzebnymi do przypisania mu świadomości. Możemy też wyobrazić sobie dwa podmioty o identycznej organizacji funkcjonalnej, które jednak w stosunku do tego samego bodźca świetlnego doświadczałyby odwróconych spektrów koloru<sup>9</sup>. Najbardziej znanym jest jednak tzw. 'argument wiedzy' (*knowledge argument*)<sup>10</sup>. Podmiot niewidomy od urodzenia nie posiada możliwości zdobycia informacji, na czym polega jakościowy charakter świadomego doświadczenia. Osoba ta, w innych aspektach normalna, wie, że istnieją doświadczenia nazywane przez ludzi kolorami. Gdyby funkcjonalizm był prawdziwy, należałoby założyć, że odpowiednio poinformowana przez osoby widzące, jak działa system funkcjonalny wzroku, mogłaby ona zrekonstruować jakościowy charakter introspekcyjnego doświadczenia barw. Albo też, gdyby przyjąć, że 'hardware' systemu wzrokowego jest czysto fizyczny, mogłaby ona uzupełnić swoją wiedzę na temat qualiów przez szczegółowe informacje naukowe na temat działania systemu wzrokowego w terminach czysto neurofizjologicznych. Wydaje się jednak, że tego rodzaju wiedza o qualiach nie mogłaby być zdobyta na żadnej z tych dwóch dróg, a stąd, że funkcjonalizm jest stanowiskiem nieadekwatnym.

Chociaż funkcjonalistyczna interpretacja umysłu nie wyklucza dualizmu, to jednak funkcjonalizm jest faktycznie głoszony wraz z teorią identyczności. Wprawdzie zawory czy pułapki na myszy nie muszą być opisywane w terminach ich materiałowych realizacji, to jednak każdy konkretny zawór jest implementowany w materiale fizycznym. W tym sensie funkcjonalizm jest teorią identyczności egzemplarycznej. Funkcjonalizm może mieć różne obszary interpretacji. W wypadku interpretacji intergatunkowej otrzymujemy mało przekonującą tezę, że psychologia, powiedzmy, człowieka, psa i kota byłaby taka sama. Można tę

<sup>9</sup> Por. N. Block, *Troubles with Functionalism*, w: C. W. Savage (wyd.), *Perception and Cognition: Issues in the Foundations of Psychology?* (Minnesota Studies in the Philosophy of Science, T 9). Minneapolis 1978, s. 261-325.

<sup>10</sup> Chodzi tu o typ argumentacji przedstawionej przez F. Jacksona w *Epiphenomenal Qualia*, „Philosophical Quarterly” 32 (1982), s. 127-136 i *What Mary Didn't Know*, „Journal of Philosophy” 83 (1986), s. 292-295. Por. też H. Robinson, *Matter and Sense*, Cambridge 1982.

tezę ograniczyć mówiąc, że redukującą teorią byłaby psychologia ludzka, natomiast teorię redukowaną stanowiłaby psychologia psów czy kotów. Oznaczałoby to, że istnieją elementy psychologii potocznej dotyczącej człowieka, którym nie odpowiadałyby żadne fenomeny w psychologii zwierząt, ale nie odwrotnie, tzn. każde prawo psychologiczne obowiązujące np. psa, obowiązywałoby również w psychologii człowieka. Wydaje się, że trudno jest dzisiaj ocenić, jak dalece tego rodzaju koncepcja jest słuszna.

W wersji introgatunkowej funkcjonalizm musi twierdzić, że zdarzenie umysłowe, np. radości, u osoby A w chwili  $t_1$  jest identyczne z pobudzeniem włókien nerwowych  $N$  w czasie  $t_1$ , lecz już w czasie  $t_2$  korelat neurofizjologiczny radości u osoby A może być inny, jak też może być różny u osoby B. Zawartości stanów umysłowych typu 'A jest przekonany, że a' lub 'B życzy sobie b', czyli zawartości tzw. postaw propozycjonalnych, stanowią przedmiot generalizacji zarówno psychologii potocznej, jak i naukowej. Z semantycznego punktu widzenia zawartości postaw propozycjonalnych są znaczeniami, tworami abstrakcyjnymi. Jako takie nie są one dostępne dla analizy czysto neurofizjologicznej. Np. wypowiedzi „Pada deszcz, a więc należy wziąć parasol” odpowiada konglomerat różnych stanów neurofizjologicznych, gdyż do przekonania 'pada deszcz' można dojść na różnych drogach fizjologicznych (przez słuch, wzrok, czucie). Za każdym razem w grę wchodzi różne systemy percepcyjne. Stąd też opis neurofizjologiczny będzie za każdym razem inny. Stanom psychicznym nie odpowiadają jednakowe stany neurofizjologiczne. Zgodnie z interpretacją funkcjonalistyczną kategorie neurofizjologiczne i kategorie psychologiczne nie przebiegają więc równoległe, lecz krzyżują się.

W latach 70. niektórzy filozofowie zaczęli się posługiwać terminem 'superweniencja' (*supervenience*) w celu wyrażenia ogólnego stanowiska materialistycznego. Kluczowym, często cytowanym tekstem, są następujące sformułowania D. Davidsona:

Chociaż stanowisko, które opisuję, przeczy istnieniu praw psychofizycznych, nie jest ono sprzeczne z poglądem, że własności mentalne są w pewnym sensie zależne od czy też superweniencjne w stosunku do własności fizycznych. Tę superweniencję można rozumieć tak oto, iż nie mogą istnieć dwa zdarzenia identyczne pod wszystkimi względami fizycznymi, ale różniące się pod względem mentalnym albo też tak, iż przedmiot nie może się zmieniać pod pewnym względem mentalnym nie podlegając żadnej zmianie fizycznej. Tego rodzaju zależność czy też superweniencja nie pociąga za sobą redukowalności przez prawo lub przez definicję; gdyby pociągała, moglibyśmy zredukować własności moralne do opisowych, a istnieją dobre racje, by wierzyć, że taka redukcja nie jest możliwa; moglibyśmy także zredukować w pewnym systemie formalnym prawdziwość do własności syntaktycznych, a wiemy, iż jest to w ogóle niemożliwe<sup>11</sup>.

Tego rodzaju stanowisko stało się atrakcyjne dla materialistycznie nastawionych filozofów, gdyż dostarczało pojęcia, które pozwalało uniknąć

<sup>11</sup> D. Davidson, *Zdarzenia mentalne*, w: B. Stanosz (wyd.), *Eseje o prawdzie, języku i umyśle*, Warszawa 1992, s.175. (D. Davidson, *Mental Events* (1970) w: D. Davidson, *Essays on Actions and Events*, Oxford 1980, s. 207-227).

konieczności akceptacji teorii identyczności lub też tezy o nomologicznej ekwiwalencji własności mentalnych i własności fizycznych. Davidson rozluźnił wymagania stawiane pod adresem pasowania do siebie różnych poziomów deskrypcji. Wydawało się, że metafizyka materialistyczna będzie mogła w ten sposób uniknąć redukcjonizmu, a więc, że ‘rozsądny’ materializm musi tylko twierdzić, iż fizyczne fakty i własności są ontycznie bazowe; że determinują wszystkie inne fakty i własności. Superweniencja wyższego rzędu własności i faktów w stosunku do faktów i własności fizycznych miała być właśnie tego typu determinacją<sup>12</sup>.

Teoria identyczności psychofizycznej może przyjmować dwie postacie: albo (jako teoria identyczności typicznej) akceptuje jednoznacznie odwracalne połączenie między stanami fizycznymi i psychicznymi (jeden stan fizyczny - jeden stan psychiczny), albo (w wersji identyczności egzemplarycznej) stwierdza możliwość połączenia wielu typów stanów fizycznych z jednym typem stanu psychicznego. Ta druga możliwość oznacza, że jeden stan psychiczny może mieć wiele fizycznych realizacji, podczas gdy odwrotna sytuacja jest wykluczona, tzn. połączenie jednego procesu neurofizjologicznego z wieloma zdarzeniami mentalnymi. To ostatnie twierdzenie wyraża istotę koncepcji superweniencji. Inaczej mówiąc, nie mogą wystąpić żadne różnice psychiczne bez przyporządkowanych im różnic fizycznych. Carrier i Mittelstra<sup>↓</sup> cytują w tym kontekście argument Puccettiego<sup>13</sup>, zgodnie z którym jakościowym zróżnicowaniem w obrębie różnych modalności zmysłowych (słuch, wzrok, smak) nie odpowiadają różnice w odpowiednich centrach słuchowych, wzrokowych i smakowych w mózgu. Nie istniałyby zatem żadne korelaty neurofizjologiczne dla jakości spostrzeżeń zmysłowych. Taki stan rzeczy empirycznie podważałby warunek superweniencji, gdyż jeden stan fizyczny byłby połączony z wieloma stanami psychicznymi. Wydaje się jednak, że jak każdy tego rodzaju argument empiryczny, także i ten stan rzeczy jest otwarty na dalsze badania, i że w związku z tym nie może być uznany za rozstrzygający.

<sup>12</sup> Por. T. Horgan, *From Supervenience to Superdupervenience: Meeting the Demands of a Material World*, „Mind” 102 (1993), s. 564-565. Termin ‘superweniencja’ pochodzi etymologicznie od łacińskich wyrazów ‘super’ (nad, nad) i ‘venire’ (przychodzić). W kontekstach filozoficznych oznacza metafizyczną (lub conceptualną) relację determinacji - to, co jest ‘superweniencje’ jest ‘ugruntowane’ przez to, na czym ‘superwieniuje’. Do analitycznej literatury filozoficznej termin ten wszedł przez klasyczną pracę z metaetyki Hare’a. Horgan cytuje następujący fragment z *The Language of Morals*: „Let me illustrate one of the most characteristic features of value-words in terms of a particular example. It is a feature sometimes described by saying that ‘good’ and other such words are names of ‘supervenient’ or ‘consequential’ properties. Suppose that a picture is hanging upon the wall and we are discussing whether it is a good picture; that is to say, we are debating whether to assent to, or dissent from, the judgement “*P* is a good picture” [...] Suppose that there is another picture next to *P* in the gallery (I will call it *Q*). [...] Now there is one thing that we cannot say; we cannot say „*P* is exactly like *Q* in all respects save this one, that *P* is good picture and *Q* not” [...] There must be some further difference between them to make one good and the other not”. (R. M. H a r e. *The Language of Morals*, Oxford 1952 s. 80-71). Na temat relacji między pojęciem ‘superweniencji’ a teorią wartości por. S. Blackburn, *Spreading the Word. Groundings in the Philosophy of Language*, Oxford 1984, s. 181-223.

<sup>13</sup> Por. R. Puccetti, *Popper and the Mind-Body Problem*, w: G. Currie, A. Musgrave (wyd.), *Popper and the Human Sciences*, Dordrecht-Boston-London 1985, s. 51-52; M. Carrier, J. Mittelstra<sup>↓</sup>, *Geist, Gehirn, Verhalten*, s. 58.

Problem umysł-ciało składa się z dwóch zasadniczych podproblemów: z właściwego zagadnienia ‘umysł-ciało’, czyli z prób odpowiedzi na pytanie o sposób połączenia umysłu z jego bazą fizyczną i ze światem (był to główny przedmiot dyskusji w tradycji filozoficznej) oraz z węższego zagadnienia, które można określić mianem zagadnienia ‘umysł - umysł’, wyrażającego się w pytaniu o charakter relacji między danym w introspekcji, świadomym umysłem a nauką (czy naukami) na jego temat. Pomijając szczegóły, można powiedzieć, że (jak dotychczas) w kwestii relacji ‘umysł-umysł’ i w kwestii natury psychologii pojawiły się cztery rozwiązania: 1. tzw. filozoficzny (logiczny) behawioryzm (w odróżnieniu od behawioryzmu naukowego lub metodologicznego, który był wyłącznie programem badawczym), zmierzający do redukcji stanów mentalnych za pomocą ich definicji w terminach dotyczących dających się obserwować bodźców i reakcji na nie; 2. program tzw. psychologii kognitywnej<sup>14</sup>, przyjmującej istnienie determinujących zachowanie, rzetelnie mentalnych elementów, nieredukowalnych do sprzężenia bodziec - reakcja, 3. podejście funkcjonalno-komputacyjne, będące metateoretyczną interpretacją psychologii kognitywnej, i polegające na hipotezie istnienia komputacyjnych relacji między stanami mentalnymi, 4. stanowisko eliminacjonistyczne, traktujące każdego rodzaju słownik mentalistyczny za przejściowy i za w dużej mierze fałszywy. Według tego ostatniego podejścia przyszłością psychologii ma być tzw. neuronauka, posługująca się wyłącznie terminami fizycznymi.

Najbardziej radykalny jest oczywiście eliminacjonizm. Przedmiotem sporu jest tutaj tzw. psychologia ludowa (potoczna) (*folk psychology*), czyli cała wiedza na temat funkcjonowania umysłu zdobyta przez ludzkość w oparciu o introspekcję<sup>15</sup>. Osądowi ze strony materializmu eliminacjonistycznego poddana jest też dotychczasowa psychologia naukowa, o ile i ona opiera się w dużej mierze na danych tzw. doświadczenia wewnętrznego. Według Churchlanda psychologia ludowa jest teorią empiryczną i jak każda tego rodzaju teoria może okazać się fałszywa<sup>16</sup>. Problemem jest jednak, w jakiej relacji znajduje się ontologia psychologii ludowej do ontologii (projektowanej przez ten rodzaj materializmu) przyszłej neuronauki. Według Churchlanda teoretycy identyczności oczekują, że psychologia ludowa zostanie ‘miętko’ zredukowana do owej pełnej neuronauki, ale jej ontologia zostanie zachowana w oparciu o transteoretyczne identyczności. Dualiści spodziewają się, że psychologia ludowa okaże się nieredukowalna ze względu na to, że jest nieredundantnym opisem autonomicznej i niefizycznej domeny fenomenów. Również i funkcjoniści liczą się z tym, że okaże się ona nieredukowalna, lecz z całkiem innych powodów

<sup>14</sup> Nie należy mylić ‘nauk kognitywnych’ (*cognitive science*) z ‘psychologią kognitywną’, która jest psychologią zorientowaną mentalistycznie, tzn. przyjmującą istnienie stanów wewnętrznych jako członów łączących bodźce i reakcje. Termin ‘psychologia kognitywna’ bywa też niekiedy używany w węższym sensie i wtedy oznacza pewne abstrakcyjne modele, mające służyć opisowi spostrzegania i myślenia. Nauki kognitywne są programem badawczym (filozofią psychologii), zmierzającym do interpretacji psychologii kognitywnej w szerokim sensie. Program ten jest oparty przede wszystkim na analogiach komputerowych, por. M. Carrier, J. Mittelstra ↓, *Geist, Gehirn, Verhalten*, s. 139.

<sup>15</sup> Na temat pojęcia psychologii ludowej (potocznej) por. zbiór artykułów w: R.J. Bogdan (wyd.), *Mind and Common Sense. Philosophical Essays on Commonsense Psychology*, Cambridge 1991.

<sup>16</sup> Por. P.M. Churchland, *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*, Cambridge MA, s. 6.

niż dualiści, tzn. sądzą, że wewnętrzna ekonomia stanów mentalnych, opisywana przez psychologię ludową, nie jest, ostatecznie patrząc, ekonomią stanów naturalnych, lecz abstrakcyjną organizacją stanów funkcjonalnych. Z tego powodu przyjmują, że stany mentalne są strukturami organizacyjnymi dającymi się urzeczywistnić w bardzo wielu rodzajach substratów, a więc że są one nieredukowalne do zasad właściwych każdemu z tych substratów. Materialiści eliminacjonistyczni są również pesymistyczni, jeśli chodzi o możliwość redukcji, lecz powodem tego pesymizmu jest przypuszczenie, że psychologia ludowa jest radykalnie nieadekwatnym wyjaśnieniem naszych wewnętrznych aktywności, wyjaśnieniem zbyt defektywnym, aby mogło pozwalać na interteoretyczną redukcję. Psychologia ludowa zostanie po prostu zastąpiona (zredukowana 'twardo') przez przyszłą neuronaukę. Churchland przedstawia następujące racje na rzecz takiej możliwości:

Jeżeli skoncentrujemy naszą uwagę nie na tym, co psychologia ludowa może wyjaśnić, lecz na tym, czego ona nie może wyjaśnić (lub na tym, czego nie jest w stanie nawet podjąć), wtedy odkryjemy, że tego rodzaju rzeczy jest bardzo dużo. Jako przykłady centralnych i doniosłych fenomenów umysłowych, które w dużej mierze (lub całkowicie) pozostają tajemnicze dla psychologii ludowej weźmy pod uwagę: naturę i dynamikę chorób umysłowych, zdolność do kreatywnej wyobraźni lub przyczyny indywidualnych różnicowań poziomu inteligencji. Rozważmy też naszą całkowitą ignorancję dotyczącą natury i funkcji psychicznych snu [...] zastanówmy się nad zwykłą zdolnością do chwycenia w biegu piłki, wylatującej za boisko lub zdolnością do trafienia przejeżdżającego samochodu kulą śnieżną. Weźmy pod uwagę wewnętrzną konstrukcję trójwymiarowego obrazu wizualnego, powstającą w oparciu o subtelne różnice w dwuwymiarowym układzie stymulacji w siatkówkach naszych oczu. Rozważmy bogatą różnorodność iluzji percepcyjnych, wizualnych i innych. Albo pomyślmy o cudzie pamięci, z jej błyskawiczną zdolnością do trafnych odtworzeń. Tych i wielu innych fenomenów psychologia ludowa nie wyjaśnia w żaden zadowalający sposób<sup>17</sup>.

Psychologia ludowa jest teorią prawdopodobnie fałszywą nie tylko dlatego, że nie potrafi wyjaśnić bardzo wielu podstawowych fenomenów umysłowych, lecz także dlatego, że nie jest heurystycznie owocna: nie zrobiła żadnych istotnych postępów od czasów Sofoklesa. Nie pasuje też do ideału teoretycznego ujednoczenia nauk w duchu fizykalizmu<sup>18</sup>. Generalizacje, którymi się posługuje, ilustrują według Churchlanda następujące przykłady: 1.  $x$  obawia się, że  $p$ , to  $x$  pragnie nie- $p$ , 2.  $x$  ma nadzieję, że  $p$  i  $x$  odkrywa, że  $p$ , to  $x$  jest zadowolony, że  $p$ , 3.  $x$  jest przekonany, że  $p$  i  $x$  jest przekonany, że jeżeli  $p$ , to  $q$ , to (wyłączając pomyłki, zaburzenia, nieuwagę itp.)  $x$  jest przekonany, że  $q$ .

W stosunku do eliminacjonizmu formułuje się zazwyczaj jeden podstawowy zarzut: jest to stanowisko 'nie do wiary'. Dokładniej rzecz biorąc, jego fałszywość wynika z jego samoznoszenia się. Gdyby był on prawdziwy, zdania opisujące jego treść należałoby uznać za zlepki śladów atramentu na papierze lub za zwykłe zakłócenia akustyczne. Napisy wyrażające materializm eliminacyjny muszą przecież coś znaczyć, a dla jakichkolwiek napisów czy dźwięków 'znaczyć' równa się posiadać odniesienie do, splecionych ze sobą na praktycznie nieskończoną ilość sposobów, stanów intencjonalnych. Asercje

<sup>17</sup> Tamże, s. 6-7.

<sup>18</sup> Por. tamże, s. 8.

czegoś (w tym oczywiście asercja materializmu eliminacjonistycznego) są pewnymi działaniami, a działania są wywoływane przez przekonania i pragnienia. Nic nie mogłoby być zatem racją dla eliminacjonizmu, skoro według niego intencjonalne stany mentalne nie istnieją (tak jak nie istnieją czarownice, a istnieją tylko neurotyczne kobiety). Racje istnieją tylko wtedy, gdy istnieją podmioty uznające te racje, a stąd też eliminacjonizm nie może nawet być racjonalnie broniony<sup>19</sup>.

---

<sup>19</sup> Por. N. Malcolm, *The Conceivability of Mechanism*, „Philosophical Review” 77 (1968), s. 45-72; A. Rosenberg. *How is Eliminative Materialism Possible?*, w: R.J. Bogdan (wyd.), *Mind and Common Sense*, s. 123-143.