Ewa Odoj
Uniwersytet Ignatianum w Krakowie
https://orcid.org/0000-0003-0821-9317
https://doi.org/10.35765/slowniki.484

Problem oddziaływania psychofizycznego

Streszczenie

DEFINICJA POJĘCIA: Problem oddziaływania psychofizycznego dotyczy pytania, jak możliwa jest interakcja pomiędzy stanami mentalnymi a fizycznymi. Rozwiązanie zależy od tego, jak pojmie się naturę umysłu.

ANALIZA HISTORYCZNA POJĘCIA: Najważniejsze historyczne stanowiska w sprawie interakcji umysł-ciało to dualizm substancjalny oraz redukcyjny fizykalizm (behawioryzm logiczny, eliminatywizm, teoria identyczności typów).

UJĘCIE PROBLEMOWE POJĘCIA: Niepowodzenie projektów o charakterze fizykalizmu redukcyjnego wynika przede wszystkim z pomijania fenomenalnego aspektu przeżyć świadomych (*qualia*). Z tego powodu najpopularniejsze dziś stanowiska dotyczące natury stanów mentalnych mają charakter nieredukcyjnego fizykalizmu. Wikłają się one jednak w zarzut epifenomenalizmu, zgodnie z którym stany mentalne, choć nie są sprowadzalne do stanów fizycznych, to jednak nie mogą oddziaływać przyczynowo w świecie fizycznym. Innym ważnym stanowiskiem jest funkcjonalizm, jednak mierzy się on z podobnymi zarzutami, co stanowiska konkurencyjne.

REFLEKSJA SYSTEMATYCZNA Z WNIOSKAMI I REKOMENDACJAMI: Współczesna refleksja nad problemem interakcji umysł-ciało powinna wziąć pod uwagę wyniki badań kognitywnych, m.in. na temat natury i genezy świadomości fenomenalnej.

Słowa kluczowe: problem psychofizyczny, koncepcje umysłu, redukcyjny fizykalizm, nieredukcyjny fizykalizm, funkcjonalizm

Określenie obszaru badań

Problem oddziaływania psychofizycznego, czyli interakcji pomiędzy umysłem a ciałem, jest centralnym zagadnieniem filozofii umysłu. Rozróżnienie na umysł i ciało ma swoje pierwotne źródło w doświadczeniu siebie jako świadomej, autonomicznej jaźni („ja"), która wydaje się niesprowadzalna do materialnego ciała. Jaźń jawi się jako podmiot stanów mentalnych, do których przynależą przykładowo takie zjawiska jak spostrzeganie, myślenie, chcenie, czucie, wyobrażanie. Pojawia się pytanie, czy i jak niematerialny i działający w sposób autonomiczny i wolny umysł może wpływać i podlegać wpływom materialnego ciała, które pozostaje pod ścisłym panowaniem praw przyrody. W tym duchu historycznie po raz pierwszy problem psychofizyczny wyraźnie postawiony został przez Kartezjusza, jako problem oddziaływania pomiędzy substancją duchową (łac. *res cogitans*) a cielesną (łac. *res extensa*), a następnie stał się jednym z centralnych zagadnień filozofii XVII wieku (m.in. dla Malebranche'a, Leibniza, Spinozy). Rozwój nauk przyrodniczych przyczynił się do podkreślenia innego aspektu problemu psychofizycznego: pytania o syntezę naszego potocznego wyobrażenia i mówienia o zjawiskach mentalnych, zwanego w filozofii umysłu psychologią potoczną (ang. *folk psychology*), z obrazem nas samych, jaki wyłania się z nauk przyrodniczych, w którym ujęci jesteśmy w sposób naturalistyczny, tzn. wyłącznie jako elementy świata przyrody. Z tej perspektywy w XX i XXI wieku, przede wszystkim w analitycznej, anglosaskiej filozofii umysłu, problem psychofizyczny nazwany został problemem umysł-ciało (ang. *mind-body problem*) lub nawet umysł-mózg (ang. *mind-brain problem*) w celu podkreślenia roli centralnego układu nerwowego w omawianym problemie.

Filozofowie umysłu starają się określić pewien zespół cech (predykatów) charakterystycznych dla stanów mentalnych i odróżniających je od stanów fizycznych (materialnych). W tradycji dualistycznej, której najważniejszym przedstawicielem jest Kartezjusz, podkreślano radykalną różnicę ontyczną umysłu i ciała, a za elementarną cechę stanów mentalnych uznawano ich nieprzestrzenność (w odróżnieniu od przestrzennej materii). Dziś wskazuje się na prywatność, intencjonalność i świadomość stanów mentalnych, choć żadna z tych własności nie jest niekwestionowalna (Bayne, 2022, s. 6–23). Prywatność (subiektywność)

stanów mentalnych oznacza, że są one dostępne poznawczo w bezpośredni, uprzywilejowany sposób wyłącznie dla „ja" doświadczającego owego stanu. Dla wszystkich innych podmiotów stany te są dostępne pośrednio (np. poprzez obserwację wyrazu twarzy osoby doświadczającej bólu lub przez opis). Takiej asymetrii nie ma w przypadku dostępu poznawczego do stanów fizycznych, ponieważ są one potencjalnie dane różnym osobom w taki sam sposób. Za Franzem Brentanem przyjmuje się, że charakterystyczną cechą stanów mentalnych, odróżniającą je od stanów fizycznych, jest ich intencjonalność, czyli skierowanie na coś (odniesienie do czegoś). Akty myślenia i sądzenia, pragnienia, emocje, spostrzeżenia itp. odnoszą się do czegoś innego od nich samych, do jakiegoś przedmiotu (niekoniecznie istniejącego realnie), zdarzenia lub stanu rzeczy. Kolejną cechą zdarzeń mentalnych jest ich świadomy charakter. Jest to z jednej strony intuicyjnie najbardziej oczywista ich własność, z drugiej – najtrudniejsza do opisu. Istnieją różne rodzaje świadomości (np. świadomość minimalna lub samoświadomość), jednak z punktu widzenia problemu psychofizycznego najważniejszy jej typ określa się jako świadomość fenomenalną. Nie sposób zdefiniować ją w sposób jednoznaczny i niecyrkularny, w związku z czym w literaturze z zakresu filozofii umysłu wskazuje się na jej intuicyjne znaczenie, używając frazy „jak to jest" (ang. *what it's like*) (Nagel, 1997). Podmioty świadome doświadczają pewnej specyficznej jakości przeżyciowej (określanej też jako *quale*, w liczbie mnogiej *qualia*), towarzyszącej ich własnym stanom świadomym, dzięki której wiedzą, jak to jest np. postrzegać kolor czerwony, jak to jest doświadczać swędzenia itp. Wiedza ta jest nieosiągalna inaczej niż poprzez własną świadomość fenomenalną (por. Hutto, 2011).

Nauki przyrodnicze – przynajmniej te bazujące na fizyce klasycznej – dostarczają nam obrazu świata materialnego jako podlegającego deterministycznym prawom przyrody (por. Penrose, 1997). Określone przyczyny z konieczności wywołują określone skutki, a każde zdarzenie w świecie materialnym ma swoje przyczynowe wyjaśnienie w innym zdarzeniu o charakterze materialnym. Tymczasem przednaukowe, ale za to elementarne intuicje psychologii potocznej podpowiadają, że mamy do czynienia z czterema rodzajami oddziaływań przyczynowych: mentalno-fizycznych, fizyczno-mentalnych, mentalno-mentalnych oraz fizyczno-fizycznych. W problemie psychofizycznym kluczową rolę odgrywa

pierwszy z nich. Doświadczamy tego, że świadome „ja" może w sposób niezdeterminowany, poprzez akty wolnej woli oddziaływać przyczynowo na przynależne mu ciało (np. poprzez podjęcie decyzji o podniesieniu ręki). Inaczej mówiąc, doświadczenie potoczne podpowiada nam, że akty mentalne mogą w sposób niezdeterminowany wywoływać pojawienie się nowego łańcucha przyczyn w świecie fizycznym. Pojawia się pytanie, w jaki sposób pogodzić ze sobą te dwie, jak się wydaje niezgodne ze sobą, perspektywy – naukową i potoczną. Problem ten wyrazić można za pomocą trylematu, czyli zbioru takich trzech twierdzeń, z których każde wydaje się prawdziwe, a równocześnie koniunkcja jakiejkolwiek pary z nich pociąga konieczność odrzucenia trzeciego twierdzenia:

1. Zjawiska mentalne różnią się od zjawisk fizycznych, tzn. nie dają się w pełni wyjaśnić w terminach zjawisk fizycznych.
2. Zjawiska mentalne oddziałują przyczynowo na zjawiska fizyczne.
3. Świat fizyczny jest przyczynowo zamknięty, tzn. każde zdarzenie w świecie fizycznym ma swoją konieczną przyczynę w innym zdarzeniu w świecie fizycznym.

Teza (3) stanowi elementarne założenie nauk przyrodniczych, a (1) i (2) zdają relację z naszego doświadczenia potocznego. (1), (2) i (3) są niekompatybilne, ponieważ: (1) i (2) implikują, że świat fizyczny nie jest przyczynowo zamknięty; (1) i (3) implikują, że co prawda zjawiska mentalne istnieją, ale nie oddziałują przyczynowo na zjawiska fizyczne; z kolei (2) i (3) implikują, że nie istnieją zjawiska mentalne różne od zjawisk fizycznych, ponieważ pierwsze redukują się do drugich. Rozwiązanie problemu psychofizycznego poprzez zanegowanie (1) i (2) charakteryzuje stanowiska materialistyczne (zwane też naturalistycznymi lub fizykalistycznymi), a odrzucenie tezy (3) proponują stanowiska dualizmu interakcyjnego (np. Kartezjusz), których specyfiką jest podkreślanie tezy (2). Z kolei za odrzuceniem (2) opowiadają się zarówno przedstawiciele dualizmu paralelnego, przyjmujący równoległość przebiegu zdarzeń mentalnych i zdarzeń fizycznych, przy równoczesnej negacji możliwości interakcji między nimi (np. Malebranche), jak i przedstawiciele współczesnych stanowisk materializmu epifenomenalnego, którzy twierdzą, że co prawda zjawiska mentalne istnieją, ale nie oddziałują na zjawiska fizyczne, choć same podlegają wpływom zjawisk fizycznych (Bremer, 2010, s. 30–32).

Analiza historyczna pojęcia

Historycznie dominującym poglądem, charakterystycznym dla epoki nowożytnej, był dualizm substancji. Zgodnie z typowym dla tego okresu założeniem, iż porządek poznania wyznacza porządek ontyczny, myśliciele nowożytni, począwszy od Kartezjusza, przyjmowali, że muszą istnieć dwa rodzaje radykalnie różniących się od siebie substancji – ciała i umysły, ponieważ w poznaniu dane są nam dwa radykalnie różne rodzaje własności – fizyczne (przede wszystkim przestrzenność) i umysłowe (przede wszystkim nieprzestrzenność). Dla dominacji stanowiska dualizmu substancji w dziejach filozofii istotne były również inspiracje religijne, ponieważ współgra ono z religijną koncepcją duszy. Za klasycznego dualistę substancji przyjmuje się Kartezjusza. Twierdził on, że człowiek jest swoistym aglomeratem ciała (materii), które ma charakter mechaniczny, oraz nieśmiertelnej i zdolnej do istnienia bez ciała duszy (umysłu).

Problem przyczynowości psychofizycznej na gruncie dualizmu substancji wydaje się szczególnie trudny do rozwiązania (albo wręcz nierozwiązywalny), ze względu na podkreślanie w nich radykalnej różnicy ontycznej umysłu i ciała oraz przypisywanie tym substancjom przeciwnych własności. Stanowiska dualistyczne, których przedstawiciele przyjmują, że choć interakcja ta jest niezwykle trudna do wyjaśnienia (albo wręcz ma charakter niewyjaśnialnej tajemnicy), to jednak jest ona faktem (zarówno w wymiarze oddziaływania mentalno-fizycznego, jak i fizyczno-mentalnego), zwane są dualizmem interakcyjnym. Próbą przezwyciężenia problemu oddziaływania psychofizycznego, przy równoczesnym zachowaniu tezy o ontycznej odrębności substancji duchowej i substancji cielesnej, jest propozycja dualizmu paralelnego. Zwolennicy tego stanowiska przyjmują albo że istnieją tylko dwa rodzaje skutecznych oddziaływań – mentalno-mentalne i fizyczno-fizyczne, albo że w ogóle nie dochodzi do jakichkolwiek bezpośrednich oddziaływań. Stanowisko to popularne było w czasach nowożytnych wśród kontynuatorów filozofii Kartezjusza. Przedstawiciele dualizmu paralelnego tego okresu zakładali, że za korelację zmian w porządku fizycznym i porządku umysłowym odpowiada Bóg, który albo te zmiany każdorazowo wywołuje (np. Malebranche), albo u początku istnienia świata przewidział wszelkie zmiany, jakie się w nim wydarzą, i ustanowił ich

harmonię (np. Leibniza). Tym samym nowożytny dualizm paralelny płacił bardzo wysoką cenę za swoją propozycję uporania się z problemem interakcji – przyjmował mianowicie bardzo mocne założenie metafizyczne, dotyczące istnienia i natury Boga, które domaga się filozoficznego uzasadnienia (Iwanicki, 2012, s. 37–41, 48–51; por. Kim, 2011, s. 31–60).

Dualizm substancji neguje fundamentalne założenie nauk przyrodniczych i z tego powodu w XX wieku został wyparty przez stanowiska fizykalizmu redukcyjnego. Niemniej jednak dualizm substancji wciąż może jawić się jako atrakcyjna koncepcja filozoficzna, ponieważ pod pewnymi względami posiada przewagę nad koncepcjami materialistycznymi o charakterze monizmu substancjalnego: lepiej radzi sobie z wyjaśnieniem doniosłego problemu tożsamości osobowej oraz pozwala na zachowanie naszych przednaukowych intuicji, dotyczących posiadania przez nas wolnej woli, dzięki której jesteśmy w stanie rozpoczynać nowe ciągi przyczynowe w świecie, wbrew determinizmowi świata przyrody. Warto zauważyć, że nie jest to kwestia błaha, ponieważ przekonanie o posiadaniu przez nas wolnej woli stanowi fundamentalne założenie życia społecznego (w tym ustroju demokratycznego), w którym przypisujemy podmiotom prawa i obowiązki.

Wśród popularnych w XX wieku stanowisk o charakterze fizykalizmu redukcyjnego wymienić należy: (1) behawioryzm logiczny, (2) teorię identyczności typów, (3) materializm eliminacyjny (eliminatywizm). Zakłada się w nich, że możliwe jest takie sprowadzenie (redukcja) stanów mentalnych do stanów fizycznych, po dokonaniu którego okaże się, że problem interakcji umysł-ciało nie jest rzeczywistym problemem, gdyż oddziaływanie, do którego się odnosi, jest po prostu oddziaływaniem tego, co fizyczne, na to, co również fizyczne. Wspólne tym stanowiskom jest więc założenie metafizycznego monizmu w kwestii ostatecznej natury rzeczywistości – istnieje tylko jeden rodzaj substancji, tzn. substancje materialne.

W latach 20. i 30. XX wieku przedstawiciele behawioryzmu logicznego (np. Gilbert Ryle) – najmocniejszej postaci fizykalizmu redukcyjnego – pod wpływem pozytywizmu logicznego Koła Wiedeńskiego twierdzili, że zjawiska umysłowe można opisać w intersubiektywnie dostępnych kategoriach behawioralnych jako dyspozycje do określonych zachowań. Postulowali oni eliminację terminu „świadomość" i wszelkich

terminów o charakterze mentalnym do terminów odnoszących się do zdarzeń obserwowalnych (np. ktoś zażywa tabletkę i mówi, że źle się czuje w przypadku odczuwania bólu). Redukcja ta miała mieć charakter konieczności analitycznej (na mocy definicji) (Heil, 2020, s. 50–70). Przedstawiciele materializmu eliminacyjnego (np. Paul Churchland), wychodząc z założenia, że tylko nauka, a ostatecznie fizyka, wyznacza to, co istnieje (tzw. realizm naukowy), głosili, że język mentalistyczny należy zastąpić językiem fizykalnym, ponieważ ten pierwszy odnosi się do czegoś, co w rzeczywistości nie istnieje. Ich zdaniem psychologia potoczna to przednaukowa prototeoria, która z czasem zostanie całkowicie wyeliminowana przez naukowy opis neurofizjologii człowieka, podobnie jak teoria spalania, jako utleniania, zastąpiła XVII-wieczną hipotezę flogistonu, który uważany był za materiał ulatniający się podczas spalania. Według przedstawicieli materializmu eliminacyjnego tak, jak hipoteza flogistonu stała się całkowicie zbędna w świetle nowej teorii fizykalnej, tak samo język mentalistyczny zostanie wyparty z naszej mowy wraz z rozwojem neuronauk (Heil, 2020, s. 166–177). Kolejną, ulepszoną wersją materializmu redukcyjnego była teoria identyczności typów (ang. *type physicalism*; *type-type identity theory*). Jej przedstawiciele (np. John Smart), zainspirowani przykładami udanych identyfikacji w przyrodoznawstwie (np. błyskawica to wyładowanie elektryczne), postulowali redukcję zdarzeń mentalnych do zdarzeń fizycznych poprzez ustalenie identyczności (tożsamości) określonych typów stanów mentalnych i określonych typów stanów neurofizykalnych, np. u wszystkich ludzi ból *K* jest tożsamy ze wzrostem aktywności włókien C_4. Redukcja ta miała mieć charakter konieczności empirycznej (na mocy prawa przyrody). Zdaniem Smarta „doznania" (stany mentalne) i „procesy mózgowe" są dwiema różnymi nazwami, które odnoszą się do tego samego przedmiotu (analogicznie do terminów „błyskawica" i „wyładowanie elektryczne"), choć jedno to wyrażenie potoczne, a drugie – wyrażenie naukowe (Heil, 2020, s. 71–87).

Programy o charakterze redukcyjnego fizykalizmu spotkały się z ostrą krytyką. Przykładowo przedstawicielom behawioryzmu logicznego zarzucano, że postulowana przez nich redukcja jest niewykonalna, ponieważ nie każde zdarzenie mentalne łączy się z jakimś zachowaniem. Możemy wyobrazić sobie istoty całkowicie odporne na ból, które pomimo doświadczania go nie reagują jakimkolwiek zachowaniem.

Podobnie moglibyśmy mieć do czynienia z osobą, która perfekcyjnie symuluje zachowanie kogoś, kto doświadcza bólu, choć w rzeczywistości go nie doświadcza. Eliminatywizm skrytykować można jako stanowisko samoznoszące: jeżeli eliminatywista jest przekonany, że nie istnieją przekonania, to tym samym podważa swój własny pogląd. Hilary Putnam (Putnam, 1975) przedstawił słynną krytykę teorii identyczności typów, zwaną zarzutem wielorakiej realizowalności (ang. *multiple realizability*). Teoria ta utożsamia określony typ zdarzeń mentalnych z dokładnie jednym typem zdarzeń fizycznych. Tymczasem jest wysoce wątpliwe, iż istnieje dokładnie jeden typ procesu neuronalnego, który związany jest np. ze wszystkimi przypadkami bólu u wszystkich istot odczuwających ból. Przykładowo tak prymitywne zwierzę, jakim jest ośmiornica, wydaje się również odczuwać ból, choć w przypadku tych istot realizuje się on na bazie innych procesów neuronalnych niż te, z jakimi mamy do czynienia w przypadku odczuwania bólu przez ludzi (Bayne, s. 38–55, por. Kim, 2011, s. 61–127).

Ujęcie problemowe pojęcia

Formułowane w XX wieku programy redukcyjnego fizykalizmu spotkały się z poważną krytyką, w wyniku której straciły na popularności. Najważniejsze zarzuty wobec stanowisk o tym charakterze odnoszą się do wspomnianego wyżej zagadnienia jakości stanów przeżyciowych (*qualiów*). Szczególne znaczenie w dyskusjach na ten temat miały opublikowane w drugiej połowie XX wieku artykuły Thomasa Nagela i Franka Jacksona. Ten ostatni w tekście *Czego Maria nie wiedziała* (Jackson, 2003) przedstawił słynny eksperyment myślowy, opisujący neuronaukowczynię Marię zamkniętą od urodzenia w czarno-białym pokoju. Posiada ona całą dostępną wiedzę na temat fizykalnych aspektów widzenia barw, ale nigdy nie widziała kolorów innych niż czarny i biały. Po wyjściu z pokoju Maria po raz pierwszy doświadczyła widzenia barw, a więc poznała fenomenalne jakości barw (*qualia*). Intuicyjnie wydaje się, że Maria zdobyła nową wiedzę, której do tej pory nie posiadała (i nie mogła posiadać), mianowicie dowiedziała się, jak to jest spostrzegać kolory. Rozumowanie Jacksona można zrekonstruować w następujący sposób:

(1) Wiedza fizykalna nie jest całą wiedzą.
(2) Jeśli fizykalizm jest prawdziwy, to wiedza fizykalna jest całą wiedzą.
(3) Fizykalizm jest fałszywy (Iwanicki, 2012, s. 66).

Eksperyment myślowy dot. Marii ma na celu wykazać prawdziwość przesłanki 1.

W podobnym duchu argumentuje Nagel w artykule *Jak to jest być nietoperzem* (Nagel, 1997). Wskazuje w nim, że nikt z nas nie wie, jak to jest orientować się w przestrzeni dzięki zdolności do echolokacji, jak czynią to nietoperze. Fizjolog może wiedzieć wszystko na temat nietoperzy, ale nigdy nie posiądzie pierwszoosobowej, doświadczalnej wiedzy na temat tego, jak to jest być nietoperzem. Zdaniem Nagela fizyczny opis świata i wszelka wiedza naukowa są zawsze niewyczerpujące, bo pomijają fenomenalny aspekt przeżyć świadomych, możliwy do poznania tylko z perspektywy pierwszoosobowej (Chalmers, 2003; Bayne, 2022, s. 133–153; Iwanicki, 2012, s. 54–63, 65–68; por. Bremer, 2005, s. 193–212).

Argumenty te zwane są argumentami z wiedzy, ponieważ broni się w nich nieredukowalnego charakteru wiedzy dot. własnych przeżyć świadomych. Równocześnie traktowane są jako mocne racje na rzecz dualizmu własności, czyli poglądu zakładającego realność i wyjątkowość zarówno własności fizycznych, jak i mentalnych. Tym samym stanowią one ważne kontrargumenty względem stanowisk o charakterze redukcyjnego fizykalizmu, którego przedstawiciele postulują redukcję własności mentalnych do własności fizycznych. W związku m.in. ze wskazaną tu krytyką projektów redukcjonistycznych popularność zyskały stanowiska klasyfikowane jako fizykalizm nieredukcyjny. Ich przedstawiciele twierdzą, że choć na sposób substancjalny istnieją jedynie ciała materialne (a zatem zakładają monizm substancjalny o charakterze materializmu), to jednak niektóre z nich posiadają obok własności fizycznych również własności mentalne, których nie da się zredukować do tych pierwszych. Tak więc na gruncie fizykalizmu nieredukcyjnego z jednej strony nie porzuca się fundamentalnego założenia przyrodoznawstwa o ostatecznej materialnej budowie świata przyrody, z drugiej natomiast nie odmawia się realności stanom mentalnym, choć równocześnie nie przypisuje się im ontycznego statusu substancji.

Nieredukcyjny fizykalizm przyjmuje trzy kluczowe tezy:

1. „Istnieją własności mentalne, które są różne od jakichkolwiek własności fizycznych".
2. „Własności mentalne zależą od (ang. *depend on*) własności fizycznych".
3. „Własności mentalne mają przyczynowy wpływ na wydarzenia [w świecie – E.O.]" (Baker, 2009, s. 110–111).

Jeśli chodzi o tezę pierwszą, to oznacza ona, iż nieredukcyjni fizykaliści przyjmują wspomniany wyżej dualizm własności. Ich zdaniem bytom materialnym mogą, choć nie muszą (jak jest w przypadku bytów nieożywionych), przysługiwać równocześnie dwa rodzaje własności – fizyczne i mentalne. Teza (2) jest sformułowana bardzo ogólnie, a jej właściwe dookreślenie stanowi najważniejsze i najtrudniejsze zadanie przedstawicieli różnych stanowisk o charakterze nieredukcyjnego fizykalizmu. Za Donaldem Davidsonem powszechnie przyjęło się określać wspomnianą w tej tezie relację mianem superweniencji (łac. *super* – na, ponad, *venire* – przybywać). Nazwa ta sugeruje nadbudowanie własności (lub zjawisk) drugiego rzędu, w tym przypadku mentalnych, na własnościach (zjawiskach) pierwszego rzędu, mających charakter podstawy – w tym przypadku fizykalnych. Zwolennicy stosowania tego pojęcia starają się zdefiniować je w taki sposób, aby oznaczało ono relację na tyle silną, by implikowała zależność stanów mentalnych od stanów fizycznych, a równocześnie na tyle słabą, aby nie pociągała redukcji jednych do drugich. Przykładowo według Davidsona zbiór własności A superweniuje na innym zbiorze B wtedy i tylko wtedy, gdy jeśli jakiś obiekt zmienił się pod względem własności mentalnych, to zmienił się również pod względem własności fizykalnych (por. Davidson, 1992, s. 175).

Teza (3) oznacza odrzucenie przez nieredukcyjnych fizykalistów epifenomenalizmu, czyli poglądu, zgodnie z którym co prawda własności mentalne istnieją i różnią się od własności fizycznych, ale są jedynie korelatami (epifenomenami) zjawisk o charakterze fizykalnym (procesów neuronalnych) i nigdy nie odgrywają roli przyczynowej w świecie fizycznym. Nieredukcyjni fizykaliści stawiają sobie tym samym za cel obronę założenia obecnego w doświadczeniu potocznym i w naukowej psychologii na temat posiadania przez stany (i zdarzenia) mentalne mocy sprawczej. Ponadto większość nieredukcyjnych fizykalistów chce pozostawać w zgodzie z naukowym obrazem świata (a właściwie dopełnić ten obraz teorią z zakresu filozofii umysłu), w związku z czym

przyjmują za niekwestionowalne założenie o przyczynowym zamknięciu świata fizycznego, tzn. założenie, iż każde zdarzenie w świecie fizycznym ma swoją wystarczającą przyczynę w innym zdarzeniu w świecie fizycznym (zob. teza (3) trylematu psychofizycznego przedstawiona powyżej) (Baker, 2009, s. 109–113; Heil, 2020, s. 178–184).

Program nieredukcyjnego fizykalizmu został poważnie skrytykowany przez Jaegwona Kima, który przedstawił między innymi tzw. argument z przedeterminowania (ang. *overdetermination*). Twierdzi w nim, że jeżeli określone zdarzenie mentalne, nazwijmy je M1, wywołuje inne zdarzenie mentalne – M2, to istnieje takie zdarzenie fizyczne F1, które jest podstawą, na której superweniuje M1, i równocześnie istnieje takie zdarzenie fizyczne F2, które jest podstawą dla M2. Jeżeli świat materialny jest przyczynowo zamknięty, co stanowi wspomniane już fundamentalne założenie nauk przyrodniczych, to F2 ma swoją wystarczającą przyczynę w F1. W związku z tym wydaje się, że M2 jest przedeterminowane, tzn. ma dwie przyczyny. Z jednej strony jest wywołane przez M1, ale równocześnie, jeśli M2 superweniuje na P2, to wystarczająca przyczyna F2, którą jest F1, jest też wystarczającą przyczyną M2. Zdaniem Kima, jeżeli zdarzenia mentalne nie są identyczne ze zdarzeniami fizycznymi, to każde zdarzenie mające przyczynę w zdarzeniu mentalnym jest przedeterminowane – ma dwie wystarczające przyczyny. Sytuację tę Kim porównuje do sytuacji, w której dwóch zabójców niezależnie od siebie postrzela śmiertelnie tę samą ofiarę w idealnie tym samym momencie. Jego zdaniem jest nieprawdopodobne, żeby tego typu sytuacja miała miejsce w każdym przypadku przyczynowania mentalnego (Baker, 2009, s. 114). Argumentacja Kima odnosi się do każdego rodzaju przyczynowania mentalnego. Powyżej przedstawiona była na przykładzie przyczynowania mentalno-mentalnego, ale z analogiczną sytuacją mamy do czynienia w odniesieniu do przyczynowania mentalno-fizycznego. Jeśli M oznacza decyzję, by podnieść rękę, P odpowiednie fizyczne zdarzenie neuronalne, które jest podstawą dla M, a E ruch ręki, to rozumowanie Kima można zrekonstruować w następujący sposób:

1. P jest przyczynowo wystarczające do wywołania E (KOMPLETNOŚĆ).
2. Jeśli P jest przyczynowo wystarczające do wywołania E, to nic, co jest inne od P, nie wywołuje E, chyba że E jest przedeterminowane (WYŁĄCZNOŚĆ).

3. M jest różne od (tzn. nie identyczne z) P (ODRĘBNOŚĆ).
4. Nie mamy do czynienia z przedeterminowaniem (NIEPRZEDETERMINOWANIE).
Więc:
5. M nie wywołuje E″ (Bayne, 2022, s. 181).

W świetle argumentacji Kima przyczynowanie mentalne wydaje się pozorne, ponieważ wszystkie zdarzenia mają swoje wystarczające przyczyny fizyczne. Jeśli rozumowanie Kima, w tym przyjęte przez niego przesłanki, są poprawne, to nieredukcyjny fizykalizm pociąga za sobą epifenomenalizm, niezależnie od deklaracji i chęci jego zwolenników. Odrzucenie tego wniosku byłoby możliwe, jeśli któraś z przesłanek nie byłaby wiarygodna. Teza o kompletności jest ściśle związana ze wspomnianym wyżej założeniem nauk przyrodniczych o przyczynowym zamknięciu świata przyrody. Teoretycznie możemy mieć do czynienia z sytuacją przedeterminowania (jak w przypadku śmierci z powodu dwóch równoczesnych postrzałów), ale takie wyjaśnienie każdego zdarzenia mentalnego wydaje się mało prawdopodobne. Przyjęcie tezy o wyłączności i nieprzedeterminowaniu wydaje się więc zasadne. Dla zachowania twierdzenia o przyczynowaniu psychofizycznym pozostaje więc odrzucić tezę o odrębności. To natomiast oznacza powrót do redukcyjnego fizykalizmu w postaci teorii, która zakłada identyczność stanów mentalnych i stanów fizycznych. Zgodnie z takim redukcjonizmem, tak jak pożar lasu jest wywołany przez wyładowanie elektryczne i piorun, ale nie jest to sytuacja przedeterminowania, bo są to dokładnie te same zjawiska, tak ruch mojej ręki jest wywołany przez fizyczne procesy neuronalne tożsame z moimi stanami mentalnymi. Wniosek z argumentacji Kima wydaje się więc następujący: albo zaakceptujemy identyczność stanów mentalnych i stanów fizycznych, albo uznamy, że nie są one identyczne, ale przyczynową sprawczość mają wyłącznie zdarzenia fizyczne. W obu przypadkach pozostaje mrzonką nasze przekonanie o możliwości wywoływania przez zjawiska mentalne (np. akty wolnej woli) nowych, niezdeterminowanych ciągów przyczynowych w świecie fizycznym (zob. teza (2) trylematu) (Bayne, 2022, s. 180–183; Kim, 2009, s. 38–46; Kim, 2011, s. 214–223; Baker, 2009, 113–116; Heil, 2020, s. 184–193).

Intensywny rozwój badań nad sztuczną inteligencją (ang. *Artificial Intelligence*, w skrócie AI) sugeruje, aby szczególną uwagę poświęcić stanowisku funkcjonalizmu, którego pierwsze sformułowania pochodzą

z połowy XX wieku. Zgodnie z podstawową ideą funkcjonalizmu określony stan umysłowy należy utożsamić z funkcją (lub rolą), jaką pełni on w odpowiednio zorganizowanym systemie. Funkcjonaliści postrzegają stany fizyczne i stany mentalne jako dwa rodzaje opisów – materialne i funkcjonalne – jednego bytu substancjalnego. Stanowisko to zrodziło się dzięki inspiracjom zaczerpniętym z informatyki, a jego źródeł doszukać się można w pracach pioniera tej dziedziny nauki – matematyka Alana Turinga. Zaproponował on rozumienie umysłów w kategoriach maszyn liczących, a stanów mentalnych jako wyników procesów obliczeniowych mózgu. Ujęcie to stanowi podstawę tzw. funkcjonalizmu maszynowego, inaczej zwanego komputacjonizmem. Stosunek między mózgiem a umysłem pojmowany jest w tej postaci funkcjonalizmu analogicznie do relacji pomiędzy komputerem (*hardware*) a jego oprogramowaniem (*software*). Zgodnie z pierwotną ideą funkcjonalizmu, zaproponowaną przez Putnama, nie jest istotne, jakie jest materialne podłoże stanów mentalnych – system, wykonujący określone funkcje umysłowe (np. mnożenie), może być zarówno organem biologicznym, jakim jest mózg, jak i układem elektronicznym. Wynika to z przyjęcia przez funkcjonalistów tezy o wielorakiej realizowalności stanów mentalnych. W związku z tym koncepcja ta dopuszcza możliwość przypisywania własności umysłowych nie tylko ludziom, ale również innym istotom żywym, a nawet urządzeniom (np. komputerom). Na gruncie komputacjonizmu mózg utożsamia się z maszyną obliczeniową, realizującą określone programy, a więc dające się wyrazić liczbowo algorytmy, które przedstawiciele tego stanowiska utożsamiają ze stanami umysłowymi. Koncepcja ta dostarczyła bardzo owocnego paradygmatu dla badań kognitywnych, w tym badań nad sztuczną inteligencją. Komputacjonizm zakłada tzw. tezę mocnej AI, zgodnie z którą maszyny liczące mogą (przynajmniej potencjalnie) myśleć, tak samo jak istoty żywe, posiadające rozbudowane biologiczne mózgi (w odróżnieniu od tzw. tezy słabej AI, zgodnie z którą maszyny mogą jedynie symulować myślenie) (Kim, 2011, s. 129–167; Van Gulick, 2009, s. 128–131; Heil, 2020, s. 88–105).

W typowej krytyce funkcjonalizmu wskazuje się m.in., że wykonywanie określonych funkcji, rozumianych jako wywoływanie tych samych skutków, co stany mentalne, nie wystarcza do tego, abyśmy mieli do czynienia ze stanami świadomymi, które w myśl tej krytyki są fundamentalną cechą stanów umysłowych. Na takie wnioski wskazują dwa

niezwykle wpływowe argumenty przedstawione w postaci eksperymentów myślowych – chińskiego pokoju autorstwa Johna Searle'a oraz chińskiego mózgu spopularyzowany przez Neda Blocka. Argumenty te podważają również tezę wielorakiej realizowalności stanów mentalnych. Wskazują bowiem, że nie każda materia może dostarczyć podłoża do pojawienia się charakterystycznych dla stanów mentalnych cech świadomości i intencjonalności.

Przeciwko tezie mocnej AI, i tym samym przeciwko komputacjonizmowi, Searle sformułował słynny argument chińskiego pokoju. Przedstawił w nim sytuację zamkniętego w pokoju człowieka, który w najmniejszym stopniu nie rozumie języka chińskiego. Z zewnątrz podawane mu są zapytania w języku chińskim, na które formułuje odpowiedź, bazując wyłącznie na książce zawierającej instrukcje, jakich znaków alfabetu chińskiego użyć w odpowiedzi na pojawienie się w zapytaniu określonego tekstu w języku chińskim. Scenariusz ten ma symulować tzw. test Turinga, którego zdanie ma według funkcjonalistów świadczyć o faktycznym myśleniu określonego systemu (maszyny Turinga). Zdaniem Searle'a osoba zamknięta w chińskim pokoju może dostarczać perfekcyjnych odpowiedzi, a mimo to w najmniejszym stopniu nie rozumieć języka chińskiego. W tej sytuacji jedynie symuluje ona świadome stany mentalne związane z rozumieniem tego języka. Ponadto, pomimo operowania znakami chińskimi, czyli syntaksą tego języka, całkowicie poza jej zasięgiem jest jej semantyka, czyli odnoszenie się znaków językowych do rzeczywistości poza nimi. Osobie tej brakuje więc tego, co nazwane zostało wcześniej intencjonalnością charakterystyczną dla stanów mentalnych.

W eksperymencie myślowym wielkiego chińskiego mózgu wskazuje się, iż na mocy zasady wielorakiej realizowalności określony mentalny stan funkcjonalny może być zrealizowany przez całą populację Chin w ten sposób, iż każdy Chińczyk wykonuje takie samo zadanie, jakie jest charakterystyczne dla pojedynczego neuronu. Każdy Chińczyk dostaje pewne dane wejściowe, na które reaguje zgodnie z posiadanymi instrukcjami, które polecają mu, jakie dane wyjściowe powinien w danej sytuacji przekazać kolejnemu Chińczykowi. W myśl tej krytyki wykonywanie takich komend przez całą populację Chin odpowiada temu, jak funkcjonaliści rozumieją stany mentalne. Tymczasem jednak nie jesteśmy w stanie wskazać, gdzie w tej złożonej z Chińczyków symulacji mózgu

mamy do czynienia ze świadomymi stanami fenomenalnymi (*qualia*). Nie można przypisać ich żadnemu pojedynczemu Chińczykowi (tak samo jak nie możemy przypisać np. świadomego postrzegania zieleni pojedynczemu neuronowi), ani tym bardziej całej populacji Chin, biorącej udział w tym eksperymencie. Wskazuje to, że funkcjonalne rozumienie stanów mentalnych jest niewystarczające, ze względu na pomijanie ich fenomenalnego charakteru (Braddon-Mitchell & Jackson 2007, s. 107–128).

Krytyka Searle'a i Blocka pozostaje do pewnego stopnia zbieżna ze wspomnianymi wyżej argumentami Jacksona i Nagela przeciwko redukcjonistycznemu materializmowi. Zarówno redukcja stanów mentalnych do stanów fizycznych, jak i redukcja do stanów funkcjonalnych (zwłaszcza rozumianych na sposób przyczynowo-skutkowy) pomija świadomościowy charakter stanów mentalnych i tym samym nie dostarcza wystarczającego opisu tego, czym one są (por. Van Gulick, 2009, s. 143–149). Ta niewystarczalność funkcjonalnej koncepcji umysłu sprawia, że zjawiska mentalne nie są na jej gruncie właściwie ujęte, a konsekwentnie uniemożliwia to udzielenie przez przedstawicieli tego stanowiska satysfakcjonującej odpowiedzi na problem interakcji psychofizycznej.

Funkcjonalizm ma jednak jeszcze inne trudności związane z problemem interakcji. Mianowicie, podobnie jak antyredukcyjny fizykalizm, funkcjonalizm wydaje się pociągać epifenomenalizm. W związku z tym, iż własności mentalne ujmowane są na gruncie tej koncepcji funkcjonalnie, a własności materialnego podłoża strukturalnie, to relacja pomiędzy tymi pierwszymi a drugimi określana jest jako realizacja (inaczej: implementacja, wykonanie). Tak jak program komputerowy realizowany jest na określonym systemie komputerowym, tak stany umysłowe realizowane są w przypadku człowieka na biologicznej materii mózgu. Własności funkcjonalne (program komputerowy, stany mentalne) stanowią tu własności drugiego rzędu, które realizują się na własnościach pierwszego rzędu – materialnych własnościach systemu (komputera, mózgu). Z tej perspektywy widać, w jaki sposób funkcjonalizm podlega analogicznej krytyce, jaką względem teorii superweniencji sformułował Kim: całość przyczynowej sprawczości przypada materialnym własnościom systemu. Realizowany stan funkcjonalny jest jedynie pochodną materialnych własności systemu, które odpowiadają za powstanie fizycznych ciągów przyczynowo-skutkowych. W związku z tym funkcjonalista

powinien albo uznać, iż przyczynowa sprawczość mentalna jest jedynie pozorna, a więc uznać stanowisko epifenomenalizmu, albo relację pomiędzy własnościami mentalnymi i fizykalnymi ująć nie tylko jako realizację określonej funkcji, ale również jako jakiś rodzaj identyczności (typów lub egzemplarzy). W obu przypadkach funkcjonaliście nie uda się rozwiązać trylematu psychofizycznego, czyli pogodzić naszych trzech intuicji. Zamiast tego w pierwszym przypadku odrzuci tezę drugą trylematu, a w drugim przypadku – pierwszą (Van Gulick, 2009, s. 142–143; Kim, 2009, s. 46–48).

Wnioski

Problem oddziaływania psychofizycznego zależy od tego, jak pojmie się stany mentalne. Jeśli zredukuje się je do stanów fizykalnych, to okaże się on pozorny, ponieważ oddziaływanie, którego dotyczy, z tej perspektywy będzie po prostu oddziaływaniem fizyczno-fizycznym. Takie ujęcie stanów mentalnych jest jednak niesatysfakcjonujące z wielu względów, choć najważniejszy z nich dotyczy pomijania świadomościowej charakterystyki stanów mentalnych, zwłaszcza przeżyciowych jakości fenomenalnych zwanych *qualiami*. Podobny problem napotykają próby zdefiniowania stanów mentalnych wyłącznie poprzez odniesienie do funkcji, jakie pełnią one w złożonym systemie. Takie ujęcia również wydają się być „ślepe" na świadomościowy wymiar stanów mentalnych – system może realizować funkcje mentalne bez typowej dla stanów mentalnych charakterystyki fenomenalnej i intencjonalnej. Jeśli natomiast przyjmie się, że stanów mentalnych nie da się zredukować, ale są to unikalne własności nadbudowane jako własności drugiego rzędu na własnościach fizycznych, to co prawda uda się zachować swoistość stanów mentalnych, ale pod poważnym znakiem zapytania stanie możliwość przyczynowej sprawczości tak pojętych stanów mentalnych. W tym ujęciu zarówno określenie relacji pomiędzy stanami mentalnymi a fizykalnymi jako relacji superweniencji, jak i jako relacji realizacji (implementacji) określonej funkcji, nie pozostawia miejsca na realną sprawczość stanów mentalnych, ponieważ zgodnie z zasadą przyczynowego zamknięcia świata fizycznego całość oddziaływania przyczynowego przypada fizycznym własnościom pierwszego rzędu.

Pomimo tych trudności zdecydowana większość badaczy zajmujących się zagadnieniami z zakresu filozofii umysłu nie podejmuje prób obrony raczej historycznego już stanowiska dualizmu substancji. Część z nich, przykładowo Daniel C. Dennett, opowiada się za epifenomenalizmem, a więc odmawia stanom mentalnym działania sprawczego. Traktują oni psychologię potoczną, w ramach której wydaje nam się, że takie oddziaływanie zachodzi, wyłącznie jako użyteczną fikcję, która pozwala przewidywać i tłumaczyć zachowania ludzi, co z kolei jest niezbędne do prowadzenia normalnego życia społecznego. Opis funkcjonalny stanów mentalnych (zwłaszcza jeśli podkreśla się w nim aspekt behawioralny) jest użyteczny z punktu widzenia rozwoju nauki i w związku z tym wciąż chętnie przyjmowany przez badaczy problematyki umysłu (nie tylko filozofów), jednak idea funkcjonalizmu zostaje zaadaptowana do pogłębiającej się wiedzy z zakresu nauk kognitywnych. Przykładowo zwraca się uwagę, że funkcjonalny poziom świadomości i intencjonalności nie pojawia się dzięki prostemu podziałowi systemu na poziom strukturalny i funkcjonalny. Wyrafinowane przejawy działania umysłu, takie jak wysoka inteligencja, intencjonalność, samoświadomość, pojawiają się dzięki integracji bardziej elementarnych podsystemów, składających się na mózg ludzki. Ideę funkcjonalizmu należy więc rozpatrywać na wielu zależnych od siebie poziomach, a nie oczekiwać pojawienia się tak zaawansowanych form funkcjonowania umysłów w pojedynczym „kroku", który stanowić miałby całe przejście od poziomu fizykalnego do mentalnego (Van Gulick, 2009, s. 138–141; por. Bremer, 2005, s. 151–192). Współcześnie ogranicza się również tezę wielorakiej realizowalności stanów mentalnych – pojawienie się stanów świadomych wydaje się ściśle związane z własnościami biologicznej materii organicznej systemów neuronalnych. Co więcej, w naukach kognitywnych pod szyldem „ucieleśnionego umysłu" podkreśla się nieodzowną rolę ludzkiego ciała w tworzeniu i funkcjonowaniu wszelkich zjawisk mentalnych. Współczesne ujęcia umysłu odcinają się więc od kartezjańskiego dualizmu substancji i elementarnych założeń tego stanowiska. Problem świadomości, zwłaszcza jej zaawansowanych form, m.in. świadomości fenomenalnej (*qualiów*), wydaje się dziś stanowić podstawową zagadkę, której rozwiązanie jest niezbędne do udzielenia odpowiedzi na problem psychofizyczny (por. Bremer, 2005).

Bibliografia

Baker, L.R. (2009). Non-Reductive Materialism W: A. Beckermann, B.P. McLaughlin, & S. Walter (Red.), *The Oxford Handbook of Philosophy of Mind* (s. 109–127). Oxford: Oxford University Press.

Bayne, T. (2022). *Philosophy of Mind*. *An Introduction*. Abingdon: Routledge.

Braddon-Mitchell, D., & Jackson, F. (2007). *The Philosophy of Mind and Cognition*. Oxford–Cambridge: Blackwell Publishing.

Bremer, J. (2005). *Jak to jest być świadomym. Analityczne teorie umysłu a problem neuronalnych podstaw świadomości*. Warszawa: Wydawnictwo IFiS PAN.

Bremer, J. (2010). *Wprowadzenie do filozofii umysłu*. Kraków: Wydawnictwo WAM.

Chalmers, D.J. (2003). Consciousness and its Place in Nature. W: S.P. Stich, & T.A. Warfield (Red.), *The Blackwell Guide to Philosophy of Mind* (s. 102–142). Malden: Blackwell Publishing.

Davidson, D. (1992). Zdarzenia mentalne. Przeł. T. Baszniak. W: D. Davidson. *Eseje o prawdzie, języku i umyśle*. B. Stanosz (Red.) (s. 163–193). Warszawa: PWN.

Heil, J. (2020). *Philosophy of Mind. A Contemporary Introduction*. New York: Routledge.

Hutto, D. (2011). *Consciousness*. W: J. Garvey (Red.), *The Bloomsbury Companion to Philosophy of Mind* (s. 35–53). London: Bloomsbury.

Iwanicki, M. (2012). Dualizm psychofizyczny. Odmiany, argumenty, zarzuty. W: M. Miłkowski, & R. Poczobut (Red.), *Przewodnik po filozofii umysłu* (s. 37–84). Kraków: Wydawnictwo WAM.

Jackson, F. (2003). Czego nie wiedziała Maria? Przeł. T. Ciecierski. *Przegląd Filozoficzno-Literacki*, *4*(6), 9–14.

Kim, J. (2009). Mental Causation. W: A. Beckermann, B.P. McLaughlin, & S. Walter (Red.), *The Oxford Handbook of Philosophy of Mind* (s. 29–52). Oxford: Oxford University Press.

Kim, J. (2011). *Philosophy of Mind*. Boulder: Westview Press.

Nagel, T. (1997). Jak to jest być nietoperzem? Przeł. A. Romaniuk. W: T. Nagel, *Pytania ostateczne* (s. 203–220). Warszawa: Fundacja Aletheia.

Penrose, R. (1997). *Makroświat, mikroświat i ludzki umysł*. Przeł. P. Amsterdamski. Warszawa: Prószyński i S-ka.

Putnam, H. (1975). Philosophy and Our Mental Life. W: H. Putnam, *Mind, Language and Reality: Philosophical Papiers.* T. 2 (s. 291–303). Cambridge: Cambridge University Press.

Van Gulick, R. (2009). Functionalism. W: A. Beckermann, B.P. McLaughlin, & S. Walter (Red.), *The Oxford Handbook of Philosophy of Mind* (s. 128–151). Oxford: Oxford University Press.