

Czy (powinniśmy uznać, że) wiedzielibyśmy wszystko o umyśle, gdybyśmy wiedzieli wszystko o mózgu?

Spór o redukcjonizm i reduktywizm w filozofii umysłu

Gwałtowny rozwój neuronauk sprawia, że stare filozoficzne pytanie o miejsce dla psychologii wśród nauk pozostaje nadal aktualne. Czy psychologia pozostanie autonomiczną dyscypliną nawet w dalekiej przyszłości (jak utrzymują antyredukcjoniści), czy też zostanie zredukowana do neuronauk (jak przewidują redukcjoniści), a jeżeli zostanie zredukowana, to czy w ogóle warto obecnie w tę dziedzinę inwestować (jak powątpiewają reduktywiści)?

Celem niniejszego tekstu jest przede wszystkim wyraźnie odróżnienie stanowiska redukcjonizmu (i antyredukcjonizmu) od stanowiska reduktywizmu (i antyreduktywizmu). Pierwsza część tekstu zarysowuje spór o redukcjonizm. Będę bronić stanowiska agnostycyzmu w tej kwestii, tj. zawieszenia sądu na temat tego, jaki będzie ostateczny kształt nauk. Najlepiej to, jaki będzie ostateczny kształt nauk zostawić naukowcom do ostatecznego wypracowania. Argumenty antyredukcjonistyczne są jednak ważne, bo zachwiewają naturalne przekonanie o jedynie możliwej słuszności redukcjonizmu. W drugiej części natomiast odniosę się krytycznie do reduktywizmu. Wykażę mianowicie, że niezależnie od tego, czy prawdziwe jest stanowisko redukcjonizmu czy antyredukcjonizmu, reduktywizm jest stanowiskiem niesłusznym, a wręcz szkodliwym.

1. Spór redukcjonizmu i antyredukcjonizmu

Redukcjonizm jest stanowiskiem spekulatywnym dotyczącym ostatecznego kształtu nauki – twierdzi mianowicie, że wszystkie nauki zostaną zredukowane do nauki podstawowej, tj. do fizyki. W szczególności psychologia (w jej ostatecznym kształcie) zostanie zredukowana do neuronauk (w ich ostatecznym kształcie). Reduktywizm to

niejako przełożenie redukcjonizmu na grunt praktyki badawczej. Według reduktywizmu należy – już teraz – preferować teorie i badania neuronaukowe nad teoriami i badaniami psychologicznymi z neuronaukami nie związanymi. Ujmując rzecz obrazowo, redukcjonizm odpowiada pozytywnie (antyredukcjonizm – negatywnie) na pytanie: „Czy *wiedzielibyśmy* wszystko o psychice gdybyśmy wiedzieli wszystko o mózgu?”. Natomiast reduktywizm odpowiada twierdząco (antyreduktywizm – przecząco) na pytanie: „Czy w naszej praktyce badawczej powinniśmy *uznać*, że *wiedzielibyśmy* wszystko o psychice, gdybyśmy wiedzieli wszystko o mózgu?”.

Na początku XX wieku panował powszechny konsensus wśród filozofów analitycznych, że wiedząc wszystko o mózgu *wiedzielibyśmy* też wszystko o psychice, a ogólniej – wiedząc wszystko o świecie mikrocząsteczek *wiedzielibyśmy* wszystko o świecie. Wszak – uważano – świat jest jeden (monizm) i jest to świat fizyczny (fizykalizm), stąd i ostateczna teoria opisująca funkcjonowanie tego świata może być tylko jedna (teza o jedności nauki), a będzie to jakaś przyszła teoria fizyczna. Oczywiście faktem jest, że istnieje nie tylko wiele teorii w samej fizyce – istnieje też wiele nauk. Ten fakt nie przeczy jednak tezie o jedności nauki – utrzymywano – albowiem teorie naukowe z «wyższych» poziomów redukują się do teorii z coraz to «niższych» poziomów aż do teorii fizycznych (redukcjonizm), tj. prawa teorii z wyższych poziomów można wyprowadzić z praw teorii z niższych poziomów (klasycznymi przykładami takich redukcji jest np. redukcja termodynamiki jakościowej do termodynamiki statystycznej, czy też redukcja genetyki klasycznej do genetyki molekularnej). W ten sposób konstruowano hierarchię nauk – z socjologią na szczycie, która miałaby się zredukować do psychologii, psychologia – do biologii, biologia – do chemii, a chemia w dużym stopniu do fizyki zredukowana już została.

Wydawać by się mogło, że stanowisko redukcjonizmu jest oczywiste. Jeżeli jednak można mówić o jakimkolwiek konsensusie wśród filozofów, to pod koniec XX wieku skłaniał się on raczej ku antyredukcjonizmowi, a więc ku przekonaniu, że nawet gdybyśmy wiedzieli wszystko o mózgu, to nie *wiedzielibyśmy* wszystkiego o psychice.

Argumentów na rzecz antyredukcjonizmu jest wiele – skoncentruję się na trzech ich klasach.

Argumenty metodologiczne

Hillary Putnam [1975a] zwraca uwagę, że to samo zjawisko można wyjaśniać na różnych poziomach, a prawa przywoływane przez wyjaśnienia z wyższego poziomu wcale nie muszą być redukowalne do praw fizyki. Aby wyjaśnić, dlaczego świerkowy kołek o kształcie walca (o określonym przekroju, którego obwód wynosi w) nie przejdzie przez kwadratowy otwór w betonowej ścianie o obwodzie w , wystarczy odwołać się do praw geometrii. Można oczywiście również odwołać się do praw mikrofizyki – wtedy trzeba potraktować zarówno kołek jak i ścianę jako strukturę atomów, obliczyć oddziałujące między nimi siły oraz obliczyć, jak siły oddziaływujące na kołek przy próbie wepchnięcia go przez otwór zmieniają układ sił wyjściowych itd. To żmudne wyjaśnienie również pozwoli nam zrozumieć, dlaczego dany drewniany kołek nie przeszedł przez dany kwadratowy otwór. Jednakże wyjaśnienie to jest pod pewnymi względami gorsze niż wyjaśnienie geometryczne. Gdyby mianowicie zmienić kołek na nieco większy, i zrobiony z dębu a nie świerka, odpowiednio powiększyć też otwór w metalowej tym razem ścianie – wówczas wyjaśnienie geometryczne nadal pozostanie to samo, natomiast nad wyjaśnieniem mikrofizycznym trzeba się będzie trudzić na nowo. Przykład ten ma nam również uzmysłowić, że wyjaśnienia z wyższego poziomu mogą być poznawczo niezależne od wyjaśnień z poziomów niższych – prawa geometrii są bądź co bądź niesprowadzalne do praw fizyki.

Po drugie, zwracano uwagę na to, że nauki z wyższych poziomów kategoryzują rzeczywistość inaczej niż nauki z poziomów niższych i dlatego trudno mieć nadzieję, że prawa nauk z wyższych poziomów zostaną zredukowane przez prawa nauk z poziomów niższych. Jerry Fodor [1974] przywołuje jako przykład prawo Kopernika-Greshama dotyczące wymiany monetarnej. Biorąc pod uwagę to, jak bardzo różne są fizyczne wcielenia pieniądza (pieniędzem mogą być kawałki różnych metali i ich stopów, kawałki papieru, także sygnały elektroniczne, a w przyszłości pieniądz może być materializowany jeszcze inaczej) trudno sobie wyobrazić, by istniały jakiegokolwiek jednorodne prawa fizyki, które rządziłyby procesami dotyczącymi pieniędzy jako takich. Nie wydaje się zatem racjonalne oczekiwać, że wszystkie prawa ekonomii zostaną zredukowane.

Po trzecie, w niektórych ujęciach filozoficznych (Cartwright [1983], Nowak [1971; 1980]), teorię naukową należy rozumieć nie jako zbiór twierdzeń (Nagel [1961]),

lecz jako hierarchiczną strukturę twierdzeń – od najprostszych praw idealizacyjnych, ujmujących zależności między najbardziej istotnymi czynnikami, do twierdzeń faktualnych (lub ich przybliżeń), ujmujących zjawiska – w miarę możliwości – w całej ich złożoności. Przy takim rozumieniu teorii pojawia się możliwość, że nawet jeżeli twierdzenia faktualne jednej teorii będą się redukować do twierdzeń faktualnych drugiej teorii, to relacja redukcji wcale nie musi zachodzić między ich prawami idealizacyjnymi – teorie te mogą uznawać inne czynniki za istotne (por. Paprzycka [2005]). W ten sposób pojawia się możliwość tego, że w jednej nauce wychwytywane są inne prawidłowości niż w drugiej nauce, mimo że obie nauki mówią o tej samej rzeczywistości.

Po czwarte, wątpliwości dotyczące redukcjonizmu powstały też na gruncie filozofii nauki, gdzie okazało się, że paradygmatyczne przykłady redukcji w nauce gwałcą wymogi stawiane redukcji. Zwracano np. uwagę na to, że Galileusza prawo wolnego spadku jest po prostu sprzeczne z Newtonowską teorią grawitacji, do której się ono rzekomo redukuje. Obok dyskusji szczegółowych pojawiały się też wątpliwości natury ogólnej, jak choćby refleksja nad tym, że pewne teorie biologiczne są ze swej natury niesprowadzalne do teorii chemicznych, jak np. teoria ewolucji. Teoria ewolucji ujmuje niejako rzeczywistość na zupełnie innym poziomie – siła naturalnej selekcji po prostu nie ma żadnego odbicia w rzeczywistości sił fizycznych, czy to na poziomie mikro, czy to na poziomie makro.

Świat jest jeden – utrzymują antyredukcjoniści – ale to nie znaczy, że aby go zrozumieć wystarczy jedna teoria, a ściślej jeden poziom rozważań.

Argument z ignorancji metafizyki umysłu

Jedną z przyczyn, dla których często przyjmuje się przekonanie o słuszności redukcjonizmu jest, po pierwsze, mylne przekonanie o tym, że w kwestii umysłu można być albo dualistą, tj. wierzyć w istnienie materialnych, śmiertelnych ciał oraz w istnienie niematerialnej, nieśmiertelnej duszy, albo redukcyjnym materialistą, tj. wierzyć, po pierwsze, że nie istnieją dusze, a jedynie ciała, a to, co zwykliśmy określać duszami czy umysłami, to po prostu nasze mózgi, a po drugie, że w związku z tym wszystko, czego możemy dowiedzieć się o naszej psychice, dowiemy się z badań nad mózgiem.

Od lat 1960tych jednym z podstawowych zadań filozofów umysłu była refleksja nad tym, w jaki sposób o relacji umysł-mózg w ogóle można myśleć właśnie na gruncie materializmu – tj. bez przywoływania istnienia substancji duchowych. Istnieją trzy podstawowe propozycje. Każda z nich ma swoje wady i zalety, a może się też okazać, że argumenty filozoficzne są po prostu niewystarczające by dokonać wyboru między nimi. W tym momencie przytaczam je jednak nie po to, by rozstrzygnąć kwestię tego, które z tych stanowisk jest słuszne, a raczej by podkreślić, rozpiętość możliwości i właśnie brak jasnego przejścia między przekonaniem o tym, że dualizm jest niesłuszny a przekonaniem o tym, że psychologia musi być zredukowana do neuronauk.

Teoria identyczności (rodzajowej). Chronologicznie pierwszym stanowiskiem była tzw. teoria identyczności (Smart [1959]), którą przedstawia się sloganem „umysł = mózg”. Zgodnie z teorią identyczności każdy stan umysłu jest tożsamy z pewnym stanem mózgu (a ściślej: każdy rodzaj stanów umysłu jest tożsamy z pewnym rodzajem stanów mózgu). Odczuwanie bólu np. jest tożsame z pobudzeniem włókien C. Dalecy jesteśmy oczywiście od odkrycia wszystkich takich tożsamości, ale rozwój neuronauk z pewnością przyczyni się do wzbogacenia naszej wiedzy w tym zakresie. Teoria identyczności (rodzajowej) jest związana z redukcjonizmem. Jeżeli jest bowiem tak, że stany umysłu są identyczne z pewnymi stanami mózgu, to znając odpowiednie tożsamości stanów umysłu i stanów mózgu oraz odpowiednie prawa neuronaukowe będziemy w stanie wyprowadzić prawa psychologiczne.

Teoria identyczności może wydawać się teorią najbardziej naturalną i oczywistą – może nawet nieść znamię naukowości – dlatego tak zaskakujący jest fakt, że jest niezgodna z pewnymi oczywistymi danymi naukowymi, a także że przesądza o tym, że program sztucznej inteligencji jest niemożliwy. Zacznijmy od drugiej kwestii. W programie sztucznej inteligencji chodzi o to, by skonstruować maszynę myślącą. Maszyna myśląca musi posiadać jakieś myśli (w tym pewną myśl M), a także dość dobrze nimi operować. Teoria identyczności mówi nam, że myśl M jest identyczna z pewnym rodzajem stanu mózgu G . To wystarczy by teoria identyczności pociągała za sobą tezę, że żadna maszyna (nie posiadająca mózgu) nie może być w stanie mózgu G , zatem nie może posiadać myśli M , a zatem nie może myśleć. Jeżeli stany umysłu są identyczne ze stanami mózgu, to tylko istoty wyposażone w mózgi mogą mieć umysły.

Sztuczna inteligencja jest niemożliwa. – Oczywiście jest wiele głębokich debat o możliwości sztucznej inteligencji, lecz taka deklaracja z filozoficznego fotela wydaje się świadczyć tylko przeciw punktowi siedzenia, a w tym wypadku przeciw teorii identyczności.

Z tego samego względu, niemożliwe byłyby jakiegokolwiek protezy, czy czipy, mające zastąpić bądź usprawnić działanie mózgu. Jeżeli pewna myśl, czy percepcja jest identyczna z pewnym stanem fragmentu mózgu, to istota, której mózg został pozbawiony tego fragmentu nie może mieć danej myśli czy percepcji. Ponownie w obliczu faktów naukowych stanowisko to wydaje się nie do utrzymania.

Funkcjonalizm. Funkcjonalizm (por. Putnam [1967; 1995]) jest stanowiskiem ratującym teorię identyczności przed powyższymi zarzutami, choć kosztem zarzucenia redukcjonizmu. Jest to stanowisko będące niejako syntezą teorii identyczności i dualizmu. Według funkjonalistów to, co sprawia, że ból jest bólem jest nie to, z jakim stanem układu nerwowego jest on identyczny, ale raczej to, jaką odgrywa funkcję – jakiego typu warunki go powodują (np. przyłożenie ręki do gorącego piecyka), jakiego typu reakcje on powoduje (np. usunięcie ręki), jaką rolę odkrywa w stosunku do innych stanów psychicznych (np. uczenia się unikania przykładania ręki do gorących przedmiotów, strach przed gorącym itd.). Funkcjonalisci chętnie przyznają, że tak rozumiany ból może być *realizowany* u ludzi (a szerzej u kręgowców) jako pobudzenie włókien C. Nie znaczy to jednak, że istota nie wyposażona we włókna C nie może odczuwać bólu – może o ile w jej organizmie istnieje pewien stan, który pełni odpowiednią rolę funkcjonalną. Funkcjonalizm stanowi też filozoficzną podstawę programu sztucznej inteligencji – maszyna może posiadać pewną myśl M o ile posiada stan, który odgrywa taką rolę funkcjonalną jak myśl M u nas. W podobny sposób możliwe są protezy układu nerwowego, o ile nie spowodują zmiany roli funkcjonalnych stanów na nich realizowanych. Slogan funkjonalizmu brzmi: „umysł ma się tak do mózgu, jak *software* do *hardware*’u”.

Wiązane z funkjonalizmem jest stanowisko teorii identyczności jednostkowej (por. Davidson [1970]), wedle której w szczególności dla stanów takich jak np. przekonania, które opatrzone są w treść, mogą nie istnieć tożsamości rodzajowe ze stanami mózgu, a jedynie tzw. tożsamości jednostkowe. Może być np. tak, że myśl o tym,

że zielone śledzie rosną na dębie przyspiewując sobie kołysankę, będzie co prawda w każdym z nas identyczna z jakimś stanem mózgu (z jakim może np. zależeć od złożonej historii rozwoju danej osoby i jej mózgu), ale wcale nie będzie tak, że w każdym z nas będzie to ten sam *rodzaj* stanu mózgu. Może być więc tak, że jedyną formą „ujrzenia” tej dziwacznej myśli będzie zidentyfikowanie jej na poziomie psychologicznym – na poziomie neurofizjologicznym będzie ona rozproszona w rozmaitych typach stanów mózgu. Funkcjonalisci przekonują, że jednoznaczne tożsamości są bardziej prawdopodobne dla prymitywnych stanów psychicznych, takich jak np. doznania bólu, ale już bardzo mało prawdopodobne dla tzw. wyższych stanów psychicznych, a w szczególności takich stanów jako opatrzone w treść przekonania, pragnienia, czy intencje.

Konsekwencją przyjęcia funkcjonalizmu jest odrzucenie redukcjonizmu (choć podważa to np. Kim [1998]; por. też Paprzycka [2005]). Jeżeli funkcjonalizm jest stanowiskiem trafnym to prawidłowości psychologiczne cechują się pewną dozą autonomii – powinny być formułowane właśnie na poziomie psychologicznym. Może być tak, że dla takich stanów jak ból np. częściowo (ale tylko częściowo) można będzie wywieść prawidłowości psychologiczne z prawidłowości neurofizjologicznych, ale takie wyprowadzenie właśnie redukcją nie będzie – będzie ono co najwyżej wyprowadzeniem prawidłowości związanych z bólem dla np. kręgowców. Z definicji jednak ponieważ ból nie jest identyfikowany z pobudzeniem włókien C, a jest co najwyżej realizowany w pobudzeniu włókien C u kręgowców, to pozostaje sfera „niewyprowadzona” z neurofizjologii kręgowców – dotycząca bezkręgowców, a może jakiś nieznanych nam jeszcze istot (np. sztucznej inteligencji). Jeszcze gorzej sprawy się mają dla redukcji ewentualnych prawidłowości dotyczących przekonań, których z zasady nie będzie można wyprowadzić z neuroprawideł – chyba że dla każdego przypadku z osobna, a to nie byłaby redukcja w zwykłym tego słowa znaczeniu.

Materializm eliminacyjny. Najbardziej radykalnym stanowiskiem jest materializm eliminacyjny (por. Churchland [1979]), który odrzuca przyjmowane przez pozostałe teorie przekonanie, że kategorie tzw. psychologii potocznej, którymi operujemy na co dzień – a więc np. kategorie przekonań, intencji, pragnień itd. – znajdują jakiegokolwiek odbicie w nauce. Według materializmu eliminacyjnego teoria psychologii potocznej jest

radykałnie złą teorią i powinna zostać po prostu odrzucona – tak jak odrzucone zostały np. flogistonowa teoria spalania, ciepłikowa teoria ciepła, czy zakłęciowa teoria chorób psychicznych. Nie ma też miejsca w ogóle dla pojęcia umysłu. Według eliminatywistów umysły nie istnieją, tak jak nie istnieje flogiston. W nauce miejsce psychologii zastąpi neuronauka. Nie ma tu zatem mowy o redukcji psychologii do neuronauk – psychologia będzie wyeliminowana.

Ten krótki przegląd trzech stanowisk materialistycznych nie wystarczy by wyrobić sobie pogląd na trafność któregośkolwiek z nich, tym niemniej wystarczy by uświadomić sobie to, że związanie przekonania o tym, że mózg jest istotny dla funkcji psychicznych wcale nie przesądza jeszcze o trafności redukcjonizmu. Każdy z omówionych poglądów przyznaje istotną rolę mózgu dla czynności psychicznych, a jednak różnią się one m.in. stanowiskiem względem losów psychologii. Według teorii identyczności przyszła psychologia zostanie zredukowana do neuronauk (a te ostatecznie do fizyki). Według funkcjonalizmu psychologia jest i pozostanie autonomiczną dyscypliną naukową, nieredukowalną do neuronauk. Według eliminatywizmu psychologia powinna zostać wyeliminowana z nauki i zastąpiona neuronauką.

Argument z paradygmatów sztucznej inteligencji

Od ponad trzydziestu lat toczy się szeroko zakrojony spór między indywidualizmem, wedle którego pojęcia psychologiczne można zdefiniować wyłącznie w odniesieniu do jednostki abstrahując od otoczenia, w którym się ona znajduje – oraz różnej maści nieindywidualizmami, wedle których pojęcia psychologiczne są w sposób istotny zakorzenione w otoczeniu poza jednostką (por. Putnam [1975b]; Burge [1979]; Wilson [1995; 2004]; Clark [1997]; Paprzycka [2002]). Nieindywidualizm jest *prima facie* sprzeczny z redukcjonizmem, a jest to stanowisko prężnie rozwijane na gruncie filozofii języka, umysłu, działania, a ostatnio także w ramach badań nad sztuczną inteligencją w paradygmacie umysłu wcielonego w świat. Dyskusje filozoficzne są zbyt specjalistyczne by je przytaczać natomiast warto przyjrzeć się trzem paradygmatom sztucznej inteligencji i temu w szczególności, jaką rolę odgrywają w nich reprezentacje. Konfrontacja tych trzech paradygmatów jest o tyle warta rozważenia, iż przyjęcie któregoś z tych paradygmatów wpływa w istotny sposób na wagę badań

neuronaukowych dla badania umysłów. Przyjmując klasyczny paradygmat – rola tych badań jest największa; przyjmując paradygmat umysłu wcielonego w świat – najmniejsza.

Systemy klasycznej SI (por. Simon [1969]) były oparte na założeniu, że umysły są systemami formalnymi, myślenie zaś polega na dokonywaniu operacji logicznych na reprezentacjach. Aby system klasyczny mógł się inteligentnie zachowywać w świecie, ów świat (bądź jego wycinek) musi mieć w systemie odpowiednie reprezentacje, na których system operuje. Stąd też bierze się jeden z najpoważniejszych problemów z programu klasycznej SI, tzw. problem ramki (por. np. Dennett [1987]), polegający po prostu na tym, że nasza wiedza zdroworozsądkowa okazuje się być trudno kodyfikowalna np. w formie przekonań – wszyscy jesteśmy przekonani, że ściany spontanicznie się nie przesuwają o 1mm/h, ale nikomu by do głowy to przekonanie nie przyszło.

Inaczej, do kwestii reprezentacji podchodzi paradygmat konekjonistyczny (por. Rummelhart, McClelland et al. [1986]). Według konekjonizmu, umysł jest ogromną siecią „neuronową,” nie ma w nim miejsca na klasyczne reprezentacje, na których system może dokonywać operacji logicznych, i do których programista ma wgląd. Systemy te używają tzw. reprezentacji rozproszonych, które są zakodowane w połączeniach między komórkami sieci. Reprezentacje te są wynikiem uczenia się systemu, mogą się różnić w zależności od m.in. architektury sieci, punktu wyjścia, etc. i są dla programistów zasadniczo nieprzejryste. Istnieją co prawda narzędzia statystyczne pozwalające na częściowy wgląd w reprezentacje rozproszone, ale w porównaniu z systemami klasycznymi reprezentacje rozproszone pozostają i tak trudno dostępne. Sprawę komplikuje fakt, że systemy konekjonistyczne ucząc się np. rozwiązywać to samo zadanie, mogą tworzyć różne konfiguracje połączeń – inaczej rozpraszając potrzebne im reprezentacje. W istocie można nabrać wątpliwości, czy w przypadku reprezentacji rozproszonych można mówić o reprezentacji świata w zwykłym tego słowa znaczeniu. A jednak systemy konekjonistyczne doskonale sobie radzą w szczególności w zadaniach, z którymi systemy klasyczne miały problemy (np. w uczeniu się czynności niealgorytmizowalnych).

Jeszcze dalej odsunięty od paradygmatu klasycznego jest tzw. paradygmat umysłu wcielonego w świat. Jednym z pionierów jest Rodney Brooks z MIT, który w swoim

klasycznym już tekście o znamienym tytule „Inteligencja bez reprezentacji” [1997] opisuje skonstruowany przez jego grupę badawczą robot o imieniu „Herbert”, którego zadanie polegało na zbieraniu puszek po napojach gazowanych w jednym z budynków w MIT. Herbert został skonstruowany na zupełnie odmiennych zasadach opierając się na myśli, że najlepszą reprezentacją świata jest sam świat. Miast powielać świat w wewnętrznych reprezentacjach, Brooks skonstruował maszynę, która potrafiła istotne dla niej informacje znajdujące się w otoczeniu odpowiednio wykorzystywać w działaniu. Oprogramowanie robota składało się z modułów od siebie niezależnych wywoływanych przez odpowiednie informacje znajdujące się w otoczeniu. Nie posiadał on też żadnego modułu centralnego, koordynującego pracę poszczególnych modułów. Robot spontanicznie poruszał się w otoczeniu, jeżeli sensory napotykały przeszkodę, uaktywniał się moduł zmiany kierunku. Kształty puszek podobne uaktywniały moduł wyciągania ramienia itd. Mimo że Herbert nie miał żadnej wewnętrznej mapy budynku, świetnie sobie radził, jeżdżąc po różnych piętrach, zbierając puste puszki. Uderzające jest to, że Herbert skonstruowany został na podstawie stosunkowo prostych zasad i mechanizmów, a wykonywał swoje zadanie świetnie.

Badania Brooksa, a także innych badaczy stanowią przykłady paradygmatu umysłu wcielonego w świat, gdzie wykracza się poza skórę/powłokę organizmu/systemu w poszukiwaniu jego umysłu. I tak Clark i Chalmers [1997] przekonują, że tak jak nie wzdragalibyśmy się przed uznaniem jakiegoś (w przyszłości) wszczepionego do mózgu modułu pamięci czy kalkulatora jako części naszego umysłu, pozwalającego nam na usprawnienie naszych funkcji poznawczych, tak nie powinniśmy się wzdragać przed uznaniem kartki papieru, kalkulatora, komputera, mapy, książki i innych narzędzi jako literalnie części naszego umysłu (o ile potrafimy się nimi sprawnie posługiwać). Nasze umysły wykraczają w istotny sposób poza naszą czaszkę, nawet poza naszą skórę.

Jeżeli pomyślimy o tych paradygmatach badań nad sztuczną inteligencją jako o paradygmatach badań nad ludzką inteligencją, to uderzające jest to jak odmienna byłaby rola przypadająca mózgowi w każdym z tych paradygmatów. Największą rolę miałby do odegrania w klasycznym paradygmacie, gdzie w jakiś sposób neuronaukowcy musieliby się doszukiwać odpowiedników reprezentacji świata i wszystkich naszych o świecie przekonań oraz odpowiedników operacji logicznych na nich dokonywanych. Trochę

skromniejszą rolę miałyby do odegrania neuronauki gdyby powielić paradygmat konecjonistyczny. Tym razem można byłoby spojrzeć na mózg jako na wielką sieć przystosowującą organizm do otoczenia za pomocą rozmaitych funkcji, ale bez tym razem narzucania programu odnajdywania klasycznych reprezentacji i operacji na nich w mózgu. Najskromniejszą rolę do odegrania miałby mózg w paradygmacie umysłu wcielonego w świat, gdzie jego rola sprowadzałaby się do odpowiedniego wykorzystania informacji znajdujących się w otoczeniu bez jej internalizowania.

Odmierna rola działalności mózgu, jaką przewidują poszczególne paradygmaty, waży też na tym, na ile możemy liczyć, że badania czysto neuronaukowe mogą nam zaoferować wyczerpującą teorię umysłu. Otóż w najlepszym razie neuronauka miałaby taką szansę gdyby właściwym paradygmatem był paradygmat klasyczny. Jeżeli jednak trafnym jest paradygmat umysłu wcielonego w świat, to po prostu nie da się dojrzeć inteligencji bez ujęcia całościowych zmagani organizmu w świecie. Tutaj rolę mózgu jest odpowiednie zgranie się z otoczeniem, wykorzystując informacje w nim zawarte bez konieczności jej powielania. Badania nad czynnościami mózgu mogą co najwyżej pomóc w zrozumieniu małej cząstki w tych zmaganiach.

Wnioski

Trudno oczywiście przypuszczać, że przytoczone argumenty na rzecz antyredukjonizmu są wystarczające by choćby wyrobić sobie w tej kwestii opinię. Błędne byłoby też mniemanie, że są to argumenty powszechnie uznane za konkluzywne. Nie było też moją intencją by to sugerować. Chodziło raczej o podważenie naturalnego przekonania o tym, że podchodząc do naukowego badania umysłu skazani jesteśmy niejako już na redukcjonizm, bo po prostu nie ma żadnych alternatyw. Otóż alternatywy są, są nawet za nimi pewne argumenty.

Wydaje się jednak, że najbardziej rozsądnym stanowiskiem – w szczególności dla osób niezaangażowanych w drobiazgową filozoficzną debatę – jest po prostu agnostycyzm, a więc zawieszenie przekonania, co do tego, czy słuszny jest redukcjonizm, czy też antyredukjonizm. Wszakże o tym, które z tych stanowisk jest słuszne moglibyśmy się dowiedzieć, gdybyśmy znali ostateczny kształt nauki, a to nie będzie nam nigdy dane.

Niemniej jednak nawet jeżeli nie jesteśmy w stanie już teraz rozstrzygnąć, która z tych alternatyw jest trafna, to ważne, że umiemy dojrzeć rozmaite możliwości. Tu praca filozofów okazuje się bezcenna, bo właśnie dzięki ich – wydawać się by mogło – czczym spekulacjom możemy otworzyć oczy szerzej. Stanowisko reduktywizmu grzeszy przede wszystkim w tym względzie, że zamyka oczy na spektrum możliwości.

2. O szkodliwości reduktywizmu

Zarówno redukcjonizm jak i antyredukcjonizm są stanowiskami spekulatywnymi – bronią wszakże tezy o tym, jaki będzie ostateczny kształt nauki. Często jednakże spotyka się zarówno wśród naukowców, jak i wśród filozofów, zwolenników stanowiska reduktywizmu. Stanowisko to jest niejako przełożeniem stanowiska redukcjonizmu na grunt praktyki badawczej. Reduktywizm przybiera wiele postaci – od ekstremalnej, wedle której powinniśmy w ogóle zarzucić np. badania psychologiczne, które nie dotyczą mózgu – do umiarkowanej, wedle której z dwóch teorii porównywalnych pod względami epistemicznymi (w podobnym stopniu sprawdzonych, w podobnym stopniu usystematyzowanych *etc.*) lepsza jest ta teoria, która jest bliższa redukcji do neuronauk.

Innymi słowy, redukcjonizm dotyczy pewnej wizji nauki ostatecznej natomiast reduktywizm wizję tę wprowadza w życie już teraz – a to, jak sądzę, jest *szkodliwe* i to niezależnie od tego, czy prawdziwe jest stanowisko redukcjonizmu czy antyredukcjonizmu. Dwa są podstawowe po temu powody.

Po pierwsze, aby w ogóle móc o redukcji mówić, muszą *istnieć* teorie z wyższego poziomu, które można byłoby redukować. Wówczas rzeczywiście redukcja takich teorii prowadzi do postępu poznawczego – zdobywamy głębszą wiedzę o tym, w jaki sposób zjawiska odkryte na wyższych poziomach zależą od zjawisk odkrywanych na niższych poziomach.

W ten to sposób – a mogłoby się wydawać paradoksalnie – *redukcjonista powinien być antyreduktywistą*, powinien dążyć do swobodnego – a więc również nieograniczonego – rozwoju nauki na różnych jej poziomach. Tylko bowiem wówczas, gdy nauki w pełni rozwiną się na wyższych poziomach można będzie próbować dokonywać redukcji.

Reduktywizm natomiast nawołuje do „redukcji na skróty” – w takim wypadku jednak wcale nie chodzi o redukcję, ale o *eliminację* nauk poziomów wyższych. To z kolei prowadzi do zubożenia naszej wiedzy, a nie jej wzbogacenia. Zilustrujmy to fikcyjnym przykładem opartym na autentycznej redukcji termodynamiki jakościowej do termodynamiki statystycznej. Zgodnie z prawem Charlesa zwiększenie temperatury gazu powoduje wzrost ciśnienia przy stałej objętości. Proces ten daje się stosunkowo łatwo wytłumaczyć w języku termodynamiki statystycznej, na gruncie której odkryto, że temperaturę gazu daje się zrozumieć jako średnią energię kinetyczną cząsteczek, natomiast ciśnienie gazu – jako średni pęd cząsteczek gazu. Z praw mechaniki można wyprowadzić zależność między wzrostem energii kinetycznej a wzrostem pędu, co wyjaśnia, dlaczego ocieplenie gazu prowadzi do wzrostu ciśnienia przy stałej objętości.

Wyobraźmy sobie fizykę, w ramach której rozwinęła się mechanika, a od kiedy odkryto świat mikrocząsteczek rozwija się głównie mechanika mikrocząsteczek, ponieważ właśnie przyjęto postawę reduktywizmu. Otóż w takiej fizyce mogłoby nigdy nie dojść do odkrycia zależności Boyle’a, Charlesa i Mariotte’a z tego prostego powodu, że gdyby patrzeć na rzeczywistość z punktu widzenia mechaniki mikrocząsteczek nikomu nie wpadłoby na myśl, żeby energią kinetyczną cząsteczek, czy pęd, w ogóle uśredniać, czy jakkolwiek wiązać z makrozjawiskami.

Po drugie, stanowisko reduktywizmu zakłóca swobodny rozwój nauki. Dopóki psychologowie będący zwolennikami redukcjonizmu stronią od badań czysto psychologicznych, koncentrując się na tym, co uznają za cenne, a mianowicie na badaniach mózgu, natomiast psychologowie będący zwolennikami antyredukcjonizmu stronią od badań na mózgu, koncentrując się na rozwijaniu psychologii – wszystko jest tak, jak być powinno. Nauka rozwija się swobodnie, a kto miał (filozoficzną) rację dowiemy się później. Martwić powinniśmy się natomiast, gdyby któraś z tych grup miała zdominować pole nauki. Na tym ucierpiałoby wszyscy.

Przykładów tego typu konfliktów podejść jest wiele, przypomnę przykład skądinąd bliski. W latach 1950tych rozwijano niezależnie dwa paradygmaty badań nad sztuczną inteligencją (SI): wspomniane już systemy klasycznej SI oraz tzw. perceptrony. Perceptrony były prekursorami współczesnych sieci neuronowych – były to systemy złożone z dwóch tylko typów komórek (wyjścia i wejścia), wyposażonych w liniowe

funkcje aktywacji. Zarówno systemy klasycznej SI, jak i perceptrony rozwijały się niezależnie, radząc sobie – każdy na swój sposób z innymi problemami. W 1969 r. Minsky i Papert opublikowali słynną krytykę perceptronów – dowiedli m. in., że mimo niewątpliwych zalet perceptrony mają ograniczone zastosowanie – w szczególności istnieje pewna klasy zadań (jak np. nauczenia się matrycy logicznej dla alternatywy wykluczonej, tzw. problem XOR), której perceptrony nie będą w stanie pokonać [1969]. Efekt tej skądinąd – i to należy podkreślić – rzetelnej krytyki był piorunujący. Oto badania nad perceptronami właściwie się zatrzymały na ponad dekadę – zatrzymały się w każdym razie fundusze skierowane na rozwój tych badań.

Jest to epizod o tyle pouczający, że fundusze te znów popłynęły, kiedy w latach 1980tych szereg badaczy pokazało, iż sieci neuronowe („perceptrony” wyposażone w dodatkową warstwę komórek, tzw. ukrytych, oraz w nieliniową funkcję aktywacji) są w stanie poradzić sobie z problemem XOR, a obiecują też radzić sobie z innymi problemami dużo lepiej niż systemy klasycznej SI.

Podkreślam – nawet jeżeli redukcjonizm jest stanowiskiem trafnym, to postawa reduktywizmu jest szkodliwa. Teorie z wyższego poziomu są ważne, gdyż to one rozpoczynają wstępne ujmowanie prawidłowości (w szczególności takich, które są ważne z praktycznego punktu widzenia), to one wypracowują metodologie badań eksperymentalnych w tej dziedzinie, wprowadzają metody pomiaru, interwencji eksperymentalnej itd. Mitem jest, że jeżeli redukcjonizm jest trafny, to równie dobrze mogłoby tych badań nie być. Gdyby tak było, to wszyscy naukowcy powinni zająć się fizyką już teraz.

Skąd się zatem bierze reduktywizm? Po części reduktywizm bierze się z nierozróżnienia tych dwóch stanowisk – spekulatywnego redukcjonizmu i praktycznego reduktywizmu – bądź też z nieświadomości sobie zagrożeń związanych z ekstrapolacją stanowiska w kwestii metafizycznej na płaszczyznę metodologiczną. Drugim źródłem popularności reduktywizmu jest fakt, że tkwimy właśnie w samym środku swoistej rewolucji naukowej, związanej nie tyle z powstaniem jakiejś nowej koncepcji, co z powstaniem nowych technologii do badania mózgu. Każdej rewolucji naukowej towarzyszy początkowo ogromna dynamika i entuzjazm badawczy (por. Kuhn [1970]). Jest to zrozumiałe, gdyż każda rewolucja naukowa otwiera – wydawać by się mogło –

nieskończone możliwości badawcze, a to przekłada się na niezwykle intensywny rozwój wiedzy.

To skądinąd zrozumiałe rewolucyjne uniesienie nie powinno jednak przeszkadzać w chłodnym osądzie. Nie wolno nam oczywiście nie doceniać już dokonanych odkryć i tych, które znajdują się na wyciągnięcie dłoni. Nie wolno nam też jednak promować badań dokonujących się tylko na jednym poziomie. Co innego być przekonanym, że ma się rację i dążyć do jej wykazania, a zupełnie co innego być przekonanym, że ma się monopol na prawdę. W nauce – jak w polityce – instytucje naukowe powinny gwarantować najdalej jak to możliwe swobodny rozwój myśli. Reduktywizm ten rozwój hamuje, przypominając o starej prawdzie – dobrymi intencjami piekło brukowane.

3. Podsumowanie

Słyszy się czasem opinię, że przecież jakoś np. przekonania muszą być w mózgu zakodowane, więc badania mózgu w końcu odpowiedź na pytanie „Jak?” muszą nam dać. Opinię tę można potraktować na przynajmniej trzy sposoby:

1. jako motywację dla badań neuronaukowych
2. jako argument za redukcjonizmem
3. jako argument za reduktywizmem

O ile badacze traktują tę opinię jako motywację do zajęcia się neuronauką, to można im tylko przyklasnąć i mieć autentyczną nadzieję, że ich rezultaty wzbogacą naszą wiedzę. Jeżeli jednak potraktować tę opinię jako argument za redukcjonizmem, a więc za przekonaniem o tym, że wszelka wiedza o umyśle jest dostępna z badań nad mózgiem, to jest to zły argument, ignorujący np. możliwość trafności funkcjonalizmu (w filozofii umysłu) czy np. paradygmatu umysłu wcielonego w świat (w sztucznej inteligencji). Jeżeli potraktować ją jako nawołanie do reduktywizmu (a więc porzucenia badań psychologicznych na rzecz neuronaukowych), to jest ona wręcz naganna, gdyż jak widzieliśmy nawet redukcjonista powinien opowiedzieć się przeciwko reduktywizmowi.

Katarzyna Paprzycka
Instytut Filozofii
Uniwersytet Warszawski
Krakowskie Przedmieście 3
00-927 Warszawa
kpaprzycka@uw.edu.pl

Literatura

- Brooks R.A. [1997]. *Intelligence without Representation*, [w:] *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, (red.) J. Haugeland, The MIT Press: Cambridge, MA, s. 395-420.
- Burge T. [1979]. *Individualism and the Mental*, [w:] *Studies in Metaphysics: Midwest Studies in Philosophy, vol. 4*, (red.) P.A. French, T.E. Uehling, Jr. i H.K. Wettstein, University of Minnesota Press: Minneapolis, s. 73-122.
- Cartwright N. [1983]. *How the Laws of Physics Lie*. Clarendon Press, Oxford.
- Churchland P.M. [1979]. *Scientific Realism and the Plasticity of Mind*. Cambridge University Press, Cambridge. W języku polskim ukazał się Rozdział 4: *Problem umysłu i ciała*, [w:] *Filozofia umysłu*, (red.) B. Chwedeńczuk, Wydawnictwo SPACJA: Warszawa, s. 47-58.
- Clark A. [1997]. *Being There: Putting Brain, Body, and World Together Again*. The MIT Press, Cambridge, MA.
- Davidson D. [1970]. *Mental Events*, [w:] *Essays on Actions and Events*, (red.) D. Davidson, Clarendon Press: Oxford, s. 207-228. Tłumaczenie polskie: *Zdarzenia mentalne*, [w:] D. Davidson, *Eseje o prawdzie, języku i umyśle*, PWN: Warszawa 1992, s. 163-193.
- Dennett D.C. [1987]. *Cognitive Wheels: The Frame Problem of AI*, [w:] *The Philosophy of Artificial Intelligence*, (red.) M.A. Boden, Oxford University Press: Oxford, s. 147-170.
- Fodor J.A. [1974]. *Special Sciences (or: The Disunity of Science as a Working Hypothesis)*. „Synthese” 28, s. 97-115.

- Kim J. [1998]. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. The MIT Press, Cambridge, MA. Tłumaczenie polskie: *Umysł w świecie fizycznym: Esej na temat problemu umysłu i ciała oraz przyczynowania mentalnego*. Wydawnictwo IFiS PAN, Warszawa 2002.
- Kuhn T. [1970]. *The Structure of Scientific Revolutions: 2nd edition*. The University of Chicago Press, Chicago. Tłumaczenie polskie: *Struktura rewolucji naukowych*. Aletheia, Warszawa 2001.
- Minsky M., Pappert S. [1969]. *Perceptrons*. The MIT Press, Cambridge, MA.
- Nagel E. [1961]. *Struktura nauki: Zagadnienia logiki wyjaśnień naukowych*. PWN, Warszawa.
- Nowak L. [1971]. *U podstaw marksowskiej metodologii nauk*. PWN, Warszawa.
- Nowak L. [1980]. *The Structure of Idealization*. Reidel, Dordrecht/Boston.
- Paprzycka K. [2002]. *False Consciousness of Intentional Psychology*. „Philosophical Psychology” 15, s. 271-295. Polska zmieniona wersja: *Falszywa świadomość psychologii intencjonalnej, czyli o pewnym dogmacie współczesnej analitycznej filozofii umysłu*. „Filozofia nauki” 2003, XI 1(41), s. 89-116.
- Paprzycka K. [2005]. *O możliwości antyredukcjonizmu*. Semper, Warszawa.
- Putnam H. [1967]. *The Nature of Mental States*, [w:] H. Putnam, *Mind, Language and Reality: Philosophical Papers*, vol. II, Cambridge University Press: Cambridge 1975, s. 429-440.
- Putnam H. [1975a]. *Philosophy and Our Mental Life*, [w:] H. Putnam, *Mind, Language and Reality: Philosophical Papers*, vol. II, Cambridge University Press: Cambridge, s. 291-303.
- Putnam H. [1975b]. *The Meaning of Meaning*, [w:] H. Putnam, *Mind, Language and Reality: Philosophical Papers*, vol. II, Cambridge University Press: Cambridge, s. 215-271. Polskie tłumaczenie: *Znaczenie wyrazu ‘znaczenie’*, [w:] *Wiele twarzy realizmu i inne eseje*, (red.) H. Putnam, Wydawnictwo Naukowe PWN: Warszawa 1998, s. 93-184.
- Putnam H. [1995]. *Umysł a ciało*, [w:] *Filozofia umysłu*, (red.) B. Chwedeńczuk, Wydawnictwo SPACJA: Warszawa, s. 185-214.

- Rumelhart D.E., McClelland J.L., the PDP Research Group , red. [1986]. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Volume 1: *Foundations*. The MIT Press, Cambridge, MA.
- Simon H.A. [1969]. *The Sciences of the Artificial*. The MIT Press, Cambridge, MA.
- Smart J.J.C. [1959]. *Sensations and Brain Processes*. „Philosophical Review” 68, s. 141-156. Polskie tłumaczenie: *Doznania a procesy mózgowo*, [w:] *Filozofia umysłu*, (red.) B. Chwedeńczuk, Wydawnictwo SPACJA: Warszawa, s. 247-262.
- Wilson R.A. [1995]. *Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind*. Cambridge University Press, Cambridge.
- Wilson R.A. [2004]. *Boundaries of the Mind: The Individual in the Fragile Sciences*. Cambridge University Press, Cambridge.