Szymon Paweł Dziczek

Uniwersytet Warmińsko-Mazurski
w Olsztynie

Warmia and Mazury University
in Olsztyn

KOMPUTACJONIZM A MATERIALIZM. CZY KRYTYKA OBLICZENIOWEJ TEORII UMYSŁU IMPLIKUJE KRYTYKĘ MATERIALNEJ JEGO NATURY?

Does the Critique of Computational Theory of Mind Imply the Negation of Its Material Nature?

S ł o w a k l u c z o w e: mózg, umysł, komputacjonizm, materializm, emergentyzm.

S t r e s z c z e n i e

Obliczeniowa teoria umysłu wywarła znaczący wpływ na kształt współczesnej filozofii zajmującej się tym fenomenem. Dopuszczenie możliwości algorytmizacji ludzkich procesów umysłowych pociąga za sobą konsekwencje zarówno w postrzeganiu umysłu pod kątem funkcjonalnym, jak i treściowym. Generuje to również pytanie o naturę tychże procesów. Odradza się tym samym znany z tradycji filozofii kontynentalnej problem relacji ciało–umysł. Odrzucenie kartezjańskiego dualizmu doprowadziło wielu przedstawicieli kognitywnego nurtu do przeświadczenia, że źródeł naszych treści mentalnych poszukiwać należy w procesach zachodzących w mózgu. Celem artykułu jest analiza argumentów adwersarzy obliczeniowej teorii, ze szczególnym uwzględnieniem ich konsekwencji dla programu wyjaśniania fenomenu umysłu w zgodzie z paradygmatem materialistycznym.

K e y w o r d s: brain, mind, computational theory, materialism, emergentism.

A b s t r a c t

The computational theory of mind has a huge impact on the entire field of the philosophy of mind. The possibility of algorithmization of human mental states has particular consequences both in a functional and semantic way. It also raises a question on the nature of those processes. The mind-body problem, known from the modern philosophy, emerges again. A rejection of Cartesian dualism led many cognitive scientists to a conviction that our mental states have their roots in the processes happening in the brain. The goal of this paper is an attempt to analyse the core arguments against the computational theory, and the consequences of its critique for the materialistic view towards the nature of the human mind.

Spór o naturę fenomenu umysłu pozostaje wciąż otwarty. Wśród rozlicznych teorii, których celem nadrzędnym jest wyjaśnienie jego pochodzenia, a przede wszystkim jego funkcji, coraz silniejszą rolę odgrywa komputacjonizm – pogląd ufundowany na przekonaniu, że umysł w swym działaniu niemal do złudzenia przypomina program komputerowy. Analogiczne mózg, będący jego substancjalną podstawą, porównywany jest z niezwykle zaawansowanym komputerem. Nietrudno się domyślić, że tak skrajne ujęcie tego fenomenu wywoływać musi stanowczy odzew w środowisku naukowym. Definiowanie umysłu przez wyjaśnienie jego warstwy funkcjonalnej zdobyło pozycję szczególną. Pytania o jego źródłowość i pochodzenie stanowią część szeroko zakrojonych prac, jednak ustępują miejsca zarówno empirycznym, jak i teoretycznym próbom wyjaśnienia jego działania. Niestety, szeroko dyskutowane rozstrzygnięcia obliczeniowej teorii obejmują najczęściej jedynie krytykę wyizolowanych tez komputacjonizmu, pomijając niejednokrotnie filozoficzne implikacje materialnego ujęcia fenomenu umysłu. W niniejszym artykule postaram się prześledzić najistotniejsze kontrargumenty kilku wybranych adwersarzy obliczeniowej teorii ze szczególnym uwzględnieniem konsekwencji dla naszego pojmowania natury umysłu w kontekście materialistycznego paradygmatu.

Materializm w filozofii umysłu jest kierunkiem niezwykle różnorodnym i trudnym do jednoznacznego zdefiniowania. W jego ramach odnaleźć możemy cały szereg rywalizujących ze sobą ujęć i teorii. Na potrzeby niniejszego artykułu najistotniejsze będą jednak dwie odmiany materializmu. Pierwszym i zarazem skrajnym jego wariantem jest materializm redukcyjny. Zakłada on możliwość redukcji stanów mentalnych (umysłowych) do zachodzących w mózgu procesów fizykalnych. Stanowisko to jest niezwykle popularne m.in. wśród naukowców zaangażowanych w projekt tworzenia sztucznej inteligencji. Drugą z postaci materializmu zwyczajowo określa się mianem nieredukcyjnego. W przekonaniu myślicieli aprobujących tę koncepcję stanów i treści mentalnych nie sposób zredukować do procesów zachodzących w mózgu. Panuje jednak wśród większości z nich zgoda, że treści te superweniują (bazują) na materialnej strukturze ludzkiego mózgu.

Jednym z najgłośniejszych krytyków obliczeniowej teorii jest amerykański filozof John Searle. W swych publikacjach odrzuca on główne założenia komputacjonizmu; *notabene* robił to już w czasach, gdy teoria ta nie uległa jeszcze krystalizacji. Jego argumenty zasiały sporo wątpliwości nawet w szeregach silnych entuzjastów „obliczeniowego" nurtu. Swoje analizy oparł Searle głównie na argumentach z dziedziny epistemologii, ontologii i językoznawstwa. Opierając się na dostępnej wiedzy o funkcjach mózgu, uznał, że doszedł do możliwego do zaakceptowania rozwiązania problemu psychofizycznego. Jak pisał w pracy *Umysł, mózg i nauka*:

> Procesy umysłowe są skutkiem działania elementów mózgu. Jednocześnie realizowane są w strukturze zbudowanej z tych właśnie elementów. Myślę, że to rozwiązanie problemu zgodne jest ze standardowym podejściem do zjawisk biologicznych. Niewątpliwie, oparty na naszej znajomości zjawisk w świecie, zdroworozsądkowy sposób rozwiązania problemu.[1]

Searle podkreśla, że takie ujęcie nie przypadło do gustu szerokiemu gronu kognitywistów, poszukujących analogii w działaniu komputera i ludzkiego umysłu. Postulat o identyczności relacji mózg–umysł i komputer–program niesie według Searle'a konsekwencje w postaci odarcia umysłu z jego biologicznej natury. Mózg zaś zostaje w ujęciu tym (silne AI) sprowadzony jedynie do roli liczącej maszyny (układu), możliwej do takiego zakodowania, by realizowany przez nią program działał identycznie z ludzkim umysłem:

> Zatem, jeśli zrobilibyśmy komputer ze starych puszek po piwie, napędzanych wiatrakami, jeśli zaprogramowalibyśmy go odpowiednio, byłby on obdarzony umysłem. Problem nie polega na tym, że w świetle naszej wiedzy taki system mógłby myśleć i odczuwać, raczej na tym, że musiałby on myśleć i czuć, gdyż jedynym warunkiem myślenia i odczuwania jest zainstalowanie odpowiedniego programu[2].

Na pytanie, dlaczego w takim razie nie udało się do tej pory stworzyć myślącego komputera, kognitywiści odpowiadają, że nie napisano jeszcze dostatecznie zaawansowanego programu i nie zbudowano dostatecznie zaawansowanego urządzenia, lecz jest to raczej kwestia czasu niż jakiejkolwiek fizycznej czy biologicznej bariery, która miałaby to uniemożliwić. Dodają również, że kolejne generacje komputerów będą posiadały umysły nie tylko równe człowiekowi, lecz w swym działaniu daleko wyprzedzające ludzkie możliwości intelektualne. Niezwykle często Searle przytacza słowa radykalnego zwolennika sztucznej inteligencji, Herberta Simona, który uważa, że komputery cyfrowe już teraz myślą w sposób równy człowiekowi. Pogląd ten wspiera jego partner w badaniach Alan Newell, dodając, że inteligencja, będąca procesem operowania symbolami, nie jest uwarunkowana wyłącznie przez biologiczną naturę mózgu. Najjaskrawszym przykładem silnej wersji sztucznej inteligencji jest jednak pogląd Marvina Minsky'ego: „będziemy szczęśliwi, jeśli maszyny zechcą zatrzymać nas w domach w charakterze domowych zwierzątek"[3].

Analizując założenia i fundamenty teorii obliczeniowej, Searle doszedł do wniosku, że jej przedstawiciele błędnie zdefiniowali pojęcie cyfrowego komputera, co przyniosło wiele niekorzystnych konsekwencji. Co więcej, by wykazać nieadekwatność komputacyjnej teorii umysłu, nie trzeba brać pod uwagę stopnia zaawansowania technologicznego. Komputer w swym działaniu, jak pisze Searle, ogranicza się do czysto formalnych operacji na symbolach (jak w przy-

---

[1] J.R. Searle, *Umysł, mózg i nauka*, PWN, Warszawa 1995, s. 25.
[2] Ibidem, s. 26.
[3] Ibidem, s. 27.

padku maszyny Turinga), które nie posiadają żadnego semantycznego znaczenia. Nie można się bowiem doszukiwać treści w sekwencjach symboli, takich jak zera i jedynki, gdyż nie reprezentują one nawet liczb rzeczywistych, odzwierciedlając jedynie formalne relacje syntaktyczne. Postulowane przez zwolenników komputacjonizmu definiowanie programu jako czysto formalnego opisu procesów zachodzących w systemie przyniosło według Searle'a możliwość falsyfikacji tez o analogii pomiędzy realizacją danego programu a procesami zachodzącymi w ludzkim umyśle. Umysł jednak, jak zauważa Searle, nie działa jedynie formalnie, gdyż jest silnie warunkowany właśnie przez treść swych stanów.

> Jeśli myślę o Kansas City, życzyłbym sobie wypić szklankę zimnego piwa bądź zastanawiam się, czy będzie spadek notowań giełdowych, w każdym wypadku mój stan umysłowy, niezależnie od tego, jakie formalne właściwości mu przypiszemy, ma jakieś psychiczne treści. To znaczy, nawet jeśli moje myśli są ciągiem symboli, musi być w myśleniu coś więcej niż abstrakcyjne symbole, gdyż ciągi symboli same w sobie nie mają żadnego znaczenia[4].

W związku z tym nie można umysłu, jak pisze Searle, przedstawić w postaci jedynie formalnej, gdyż każdy jego stan niesie za sobą pewną treść. Co za tym idzie – myślenie nigdy nie jest bezprzedmiotowe, a tym samym szeregu formalnych znaków za myśl uznać nie sposób. Wysunął w ten sposób kolejny poważny zarzut wobec obliczeniowej teorii, twierdząc, że formalny program komputerowy, pozbawiony z definicji treści, z pewnością nie posiada identycznego charakteru co ludzki umysł[5]. Treściowy (semantyczny) aspekt umysłu jest przez Searle'a mocno eksponowany, w przeciwieństwie do badaczy nurtu komputacyjnego, którzy w swych pracach silnie go deprecjonowali. Zasadnicza argumentacja Searle'a oparta jest na rozważaniach filozoficznych i logicznych, w mniejszym zaś stopniu skupia się na zasadności komputacjonizmu z punktu widzenia matematyki czy informatyki.

Swój naczelny argument skierowany przeciwko obliczeniowej teorii umysłu zobrazował Searle za pomocą słynnego teoretycznego eksperymentu zwanego chińskim pokojem. Zadaje w nim pytanie: czy jeśli cyfrowy komputer, wyposażony w program operujący językiem dla przykładu chińskim, będzie w stanie odpowiadać na zadawane mu pytania równie dobrze, jak mieszkaniec tego kraju, to czy można powiedzieć o komputerze, że zrozumiał język chiński? Analogiczna do tego byłaby według Searle'a sytuacja, gdyby człowieka, który nie zna języka chińskiego, umieścić w zamkniętym pokoju i dostarczyć mu zbiór chińskich znaków i symboli oraz poradnik napisany w rozumianym przez niego języku, traktujący o tym, jak tymi znakami operować. Poradnik opisuje jednak, jak formułować konstrukcje znaków jedynie od strony formalnej, syntaktycznej. Do pokoju napływają ciągi symboli; osoba zaś na podstawie reguł składa ze znaków

[4] Ibidem, s. 28.
[5] Ibidem.

inne konstrukcje. W ten sposób, nie będąc tego świadoma, prowadzi rozmowę z ludźmi na zewnątrz, a poradnik, który otrzymała, jest tak zaawansowany, że nie możliwe jest odróżnienie jej od osoby doskonale znającej język chiński. Według Searle'a, o rozumieniu języka w tym przypadku nie może być mowy:

> Realizując taki formalny komputerowy program, z punktu widzenia operatora z zewnątrz, zachowujemy się dokładnie tak, jak byśmy rozumieli język chiński, jednocześnie jednak nie znamy ani jednego słowa z języka naturalnego. Jeśli wykonanie komputerowego programu symulującego wykonanie języka chińskiego nie jest wystarczające dla nas, byśmy ten język rozumieli, nie może być wystarczające także dla maszyny liczącej[6].

Gdy człowiekowi z chińskiego pokoju zadać pytanie w jego rodzimym języku, z łatwością odpowie właśnie dlatego, że symbolom z języka przypisane są konkretne znaczenia semantyczne, które natychmiast odkrywają przed nim sens wypowiedzi.

Eksperyment myślowy Searle'a dowodzi (w jego mniemaniu), że formalne reguły i procedury nie pozwalają człowiekowi na zrozumienie i nauczenie się języka, a co za tym idzie – nie może tej umiejętności posiąść również maszyna. Co więcej, budowa cyfrowego komputera uniemożliwia mu poznanie semantycznej treści, może on jedynie posiadać reguły formalne, protokoły działania, a w wypadku języka – jedynie aspekt syntaktyczny, nigdy zaś semantyczny. Eksperyment chińskiego pokoju ma też szersze konsekwencje. Przekonuje, że nasze stany myślowe, niezależnie, czy dotyczą operowania językiem, czy jakiejkolwiek innej działalności, nigdy nie opierają się jedynie na formalnej manipulacji symbolami czy znakami; silnie za to nacechowane są treścią. Komputery cyfrowe zaś, w całej swej złożoności, niezależnie od poziomu zaawansowania technologicznego tych treści nie posiadają, stąd odpowiedź na pytanie o analogię pomiędzy myśleniem komputerów i ludzi musi być przecząca.

> Nie ma takiego programu komputerowego, który sam w sobie wyposażałby system w umysł. Mówiąc krótko, programy nie są umysłami ani same w sobie nie wystarczą dla powstania umysłu. [...] Czynności mózgu ograniczone tylko do realizowania programu komputerowego nie wystarczą, by funkcjonowanie mózgu doprowadziło do powstania umysłu[7].

Jeśli więc umysł jest właściwością ludzkiego mózgu, a obsługiwanie przez mózg tylko formalnego programu nie prowadzi do powstania stanów umysłowych, nie doprowadzi również w konsekwencji do powstania samego umysłu. Należy zatem według Searle'a uznać, że umysł posiada większą złożoność niż zwolennicy obliczeniowej teorii wcześniej założyli. Istotnym również aspektem mózgu jest jego biologiczna natura, od której nie sposób uciec, analizując właściwości ludzkiego umysłu. Można by rzec, że charakter funkcjonowania organicznego umysłu jest bezpośrednio uwarunkowany przez tę biologiczność:

---

[6] Ibidem, s. 29.

[7] Ibidem, s. 35–36.

„Cokolwiek, co mogłoby być przyczyną umysłu, musiałoby mieć moc oddziaływania przyczynowego porównywalną z możliwościami mózgu”[8].

Ogromna złożoność ludzkiego mózgu, którego aktywność jest bezpośrednią przyczyną powstania umysłu, wymaga od jakiegokolwiek innego systemu równie wysokiego stopnia złożoności, by procesy przez niego realizowane powodowały stany umysłowe. Komputery cyfrowe, z zasady pozbawione semantyki, nie zaliczają się jednak do tego typu układów, stąd w środowisku sprzętowym powstanie umysłu jest z definicji niemożliwe:

> Sądzę – pisze Searle – że ostateczny wynik dyskusji przypomniał nam coś, co już od dawna wiemy, że stany umysłowe są zjawiskiem biologicznym. Świadomość, intencjonalność, subiektywność, moc przyczynowego oddziaływania umysłu, wszystko to należy do dziejów biologicznego życia, razem ze wzrostem, rozmnażaniem, wydzielaniem żółci i trawieniem.[9]

John Searle w swych wieloletnich rozważaniach nad naturą umysłu za cel nadrzędny postawił sobie wypracowanie teorii akcentującej jego związek z biologiczną naturą mózgu. Próba zakwestionowania kartezjańskiego dualizmu towarzyszyła w jego filozofii krytyce radykalnego materializmu. Najsilniej zaś polemizował ze stanowiskiem epifenomenalizmu, zakładającego w odmianie skrajnej możliwość redukcji treści mentalnych bezpośrednio do fizykalnych stanów mózgu bądź w wersji umiarkowanej, że pewne stany fizjologiczne posiadają również komponentę mentalną:

> Cóż można odpowiedzieć na epifenomenalizm? Jedna uwaga nasuwa się natychmiast: byłoby to niezwykłe, niepodobne do niczego, co kiedykolwiek zdarzyło się w historii przyrody, gdyby istnienie jakiegoś tworu biologicznego, tak złożonego, bogatego i rozbudowanego jak ludzka i zwierzęca świadomość, w żaden sposób nie wpływało na świat realny. Zgodnie z tym, co wiemy o ewolucji, mało prawdopodobne jest, żeby epifenomenalizm był poglądem słusznym. Nie daje to podstaw do ostatecznego odrzucenia epifenomenalizmu, ale powinno przynajmniej sprawić, że skrzywimy się na myśl o nim.[10]

W konsekwencji systematyzacja i klasyfikacja poglądów Searle’a pozostawała przez wiele lat problematyczna. Należy wziąć pod uwagę kilka zasadniczych jego rozstrzygnięć:

- odrzucenie materializmu redukcyjnego,
- odrzucenie kartezjańskiego dualizmu,
- podkreślenie biologicznej natury mózgu i jego związek z umysłem,
- uznanie silnej roli sprawczej ludzkiej intencjonalności.

Z analizy powyższych założeń wysnuć można wniosek o pewnej formie emergentyzmu. Istotna dla rozważań jest charakterystyczna dla tego nurtu teza o możliwości wyłonienia się nowych, niespotykanych dotychczas dla systemu

---

[8] Ibidem, s. 36.

[9] Ibidem.

[10] J.R. Searle, *Umysł, język, społeczeństwo*, W.A.B., Warszawa 1999, s. 98.

cech na wyższych poziomach złożoności. Co prawda na pytanie o poziom, na którym zarysowuje się fenomen umysłu (neuronalny czy nawet niższy), Searle nie podaje odpowiedzi, jednak nie ma on wątpliwości, że przyczyną jego powstania są procesy zachodzące w mózgu. Co więcej, poczynione przez niego rozstrzygnięcia dopuszczają swoistą formę superweniencji kauzalnej. Termin ten wydaje się odpowiedni o tyle, że wskazuje na kluczową dla Searle'a rolę przyczynowej sprawności treści mentalnych na fizjologiczne zmiany zachodzące w naszym ciele. Nie zgadza się on z zasadą przyczynowego domknięcia sfery fizycznej, wedle której zjawiska w świecie fizycznym są przyczynowo niezależne od jakichkolwiek treści umysłowych. Tę opozycję Searle lubi wyrażać często przytaczanym w czasie wykładów przykładem: „I decide to raise my arm and the damn thing goes up" (Postanawiam podnieść rękę i ta cholerna rzecz się podnosi)[11].

Reasumując, najświeższe wystąpienia tego amerykańskiego filozofa i jego dotychczasowe naukowe publikacje wskazują, że nie odrzuca on umiarkowanej bądź „słabej" wersji materializmu. Oczywiście negowaje materializm w wersji redukcyjnej, a także wszelkie implikacje epifenomenalizmu i tym silniej kartezjański dualizm, który uważa za jedną z najbardziej niefortunnych idei filozofii kontynentalnej. W jego ujęciu krytyka obliczeniowej teorii umysłu nie pociąga jednak za sobą zakwestionowania jego materialnego fundamentu, gdyż jego źródeł upatrywać należy w biologicznej strukturze naszego mózgu.

W 1931 r. austriacki logik i matematyk Kurt Gödel sformułował dwa twierdzenia: (1) o niezupełności i (2) niedowodliwości niesprzeczności[12]. Odnoszą się one do opartych na arytmetyce aksjomatycznych systemów matematycznych: „Dla każdego systemu sformalizowanego matematyki i dostatecznie bogatego o obliczalnym (rekurencyjnym) zbiorze aksjomatów i obliczalnych relacjach inferencji oraz niesprzecznego istnieje zdanie zbudowane w języku tego systemu, które jest w nim nierozstrzygalne, zdanie to nazywane jest formułą Gödlowską"[13]. Drugie z twierdzeń dotyczy braku możliwości udowodnienia niesprzeczności pewnego układu, opierając się jedynie na metodach, narzędziach tego układu: „Istnieje takie prawdziwe zdanie systemu o liczbach naturalnych, którego prawdziwości nie można udowodnić w ramach tego systemu"[14].

Twierdzenia Gödla przyniosły poważne matematyczne konsekwencje. Jednak istotne dla rozważań nad komputacyjną teorią umysłu jest to, jak wpłynęły na rozumienie procesu dowodzenia. Jak twierdził Gödel, niemożliwa jest algoryt-

---

[11] Wykład w University of Cambridge, 22 maja 2014, Consciousness as a Problem in Philosophy and Neurobiology. Zob. [online] <www.youtube.com/watch?v=6nTQnvGxEXw>.

[12] K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, t. I*, „Monatshefte für Mathematik und Physik" 1931, nr 38, s. 173–198.

[13] M. Hetmański, *Umysł a maszyny. Krytyka obliczeniowej teorii umysłu*, Wydawnictwo UMCS, Lublin 2000, s. 149.

[14] Ibidem, s. 149–150.

mizacja wszystkich dowodów matematycznych, stąd operacje dowodzenia wykonywane zarówno przez maszynę, jak i przez człowieka mają zgoła odmienny charakter. Wykazując zasadnicze granice algorytmizacji i mechanizacji, przekonywał w konsekwencji o odmienności natury procesów umysłowych człowieka i formalnego obliczeniowego dowodzenia.

Według Gödla, ludzkiego umysłu nie można więc nazywać maszyną, gdyż w jego działaniu wyróżnić można zasadniczo niemechaniczne operacje wnioskowania, w przeciwieństwie do formalnych, mechanicznych procedur, które to już od czasów maszyny Turinga są fundamentem obliczeniowej teorii. „Rozwój umysłu – przynajmniej w odniesieniu do pracy matematyka – to głównie doskonalenie czynności wnioskowania (również poszukiwanie nowych), które może się dokonywać również niemechanicznie i być poza jakąkolwiek fizyczną (maszynową) realizacją, jako proces intuicyjnego wglądu w istotę prawdy. Nie ma on wyłącznie czy zasadniczo czysto mechanicznego charakteru. Intuicyjna działalność umysłu polega na odkrywaniu (a nie konstruowaniu) obiektywnych prawd matematycznych. Intuicji tej nie można zatem do końca zmechanizować”[15]. Ograniczenia mechaniczności nie zamykają jednak matematykom drogi do sprawdzania prawdziwości zdań pozasystemowych innymi niż formalne metodami. „Pojęcia prawdziwości semantycznej nie można więc adekwatnie wyrazić w terminach prawdziwości syntaktycznej tzn. dowodliwości”[16].

Rozważając wysuwane przez Gödla argumenty na rzecz odróżnienia natury operacji wykonywanych przez maszynę i myślącego człowieka, dostrzec można wizję umysłu zgoła odmienną od dzisiejszego paradygmatu, opracowanego w obrębie neuronauk i kongitywistyki. Gödel dopuszcza bowiem niemechaniczne i nierealizowalne fizycznie operacje, jak choćby zacytowaną wyżej możliwość intuicyjnego wglądu w istotę rzeczy. Wydaje się, że odrzucenie możliwości fizycznej realizacji procesów niemechanicznych przekreśla z konieczności materialną ich źródłowość.

Jak podkreślał sam Gödel – „materializm jest fałszywy”[17]. W przypadku austriackiego matematyka, zakwestionowaniu obliczeniowej teorii towarzyszy również sceptycyzm wobec materializmu jako takiego. Inaczej niż Searle, Gödel zakłada istnienie w człowieku również sfery niedefiniowalnej za pomocą fizykalnych kategorii i terminów, jak również jest przekonany o istnieniu niematerialnej duszy i niematerialnych komponentów ludzkiego umysłu[18].

W konsekwencji odrzucony musi zostać w tym ujęciu nie tylko materializm w wersji redukcyjnej. Również nieredukcyjne teorie relacji pomiędzy sferami fi-

[15] Ibidem, s. 151.

[16] R. Murawski, *Funkcje rekurencyjne i elementy metamatematyki*, UAM, Poznań 1990, s. 159.

[17] Hao Wang, *A logical journey. From Gödel to philosophy*, MIT Press, Cambrige 1996.

[18] Hao Wang, *From mathematics to philosophy (International library of philosophy and scientific method)*, Humanities Pres 1974.

zyczną i mentalną zostają zakwestionowane. Pod uwagę wziąć należy fakt, że za życia Gödla zagadnienie superweniencji w dzisiejszym jej rozumieniu w obszarze filozofii umysłu jeszcze nie występowało. Jednak poczynione przez niego rozstrzygnięcia podsuwają dostateczne przesłanki do próby klasyfikacji jego koncepcji. Z całą pewnością możliwe jest wykazanie rozłączności wyżej wymienionych teorii, opierając się na analizach filozoficznych i metodologicznych konsekwencjach osiągnięć naukowych austriackiego matematyka.

Jako pierwszy filozoficznymi implikacjami prac Gödla zajął się brytyjski filozof John Lucas w swojej pracy *Minds, machines and Gödel* (1961). Szczególnie ciekawe było uwzględnienie w niej Gödlowskich twierdzeń w odniesieniu do komputacyjnej wizji umysłu. Lucas analizował również z dużym zaangażowaniem osiągnięcia Alana Turinga i polemizował z prezentowanymi przez niego teoriami. Głównym założeniem rozważań Lucasa było kategoryczne rozróżnienie struktury ludzkiego umysłu i budowy maszyny Turinga, oparte na twierdzeniach Gödla, które uznał za wystarczający dowód dla ich systematyzacji. Według Lucasa, niektóre realizowane przez człowieka procesy poznawcze z natury nie mogą być przez maszynę symulowane. Co więcej, odrębny charakter tych działań wskazuje na zasadniczą nadrzędność ludzkiego umysłu względem funkcjonowania maszyny: „Wobec maszyny Turinga umysł ludzki ma bowiem tę zasadniczą przewagę, że potrafi sformułować zarówno samą treść, jak i uświadomić sobie epistemologiczny sens twierdzenia Gödlowskiego [...], czego żadna maszyna nie jest w stanie dokonać”[19].

Jak twierdzi Lucas, umiejętność ta, właściwa człowiekowi, jest ważną częścią procesu poznania. Nie posiada jej za to maszyna, dlatego odrzucić należy możliwość symulowania przez nią tych stanów oraz wysuwany przez zwolenników komputacjonizmu postulat o identycznej strukturze umysłu i maszyny liczącej. W oparciu o twierdzenie Gödla postuluje Lucas, że zakres działań umysłu jest znacznie szerszy niż chcieliby tego stronnicy kognitywnego nurtu:

> Tym samym nie możemy mieć nadziei na wytworzenie maszyny, która będzie w stanie robić wszystko, co może robić umysł; nie możemy nigdy mieć, nawet w zasadzie, mechanicznego modelu umysłu.[20]

Ograniczenia strukturalne maszyny uniemożliwiają jej analizę stanów, w których się znajduje, bez wyjścia poza system ją kształtujący. Aby jednak odpowiedzieć na pytanie o prawdziwość pewnego zdania o systemie, niezbędne wydaje się wkroczenie na poziom wyższy niż formalnie zaprogramowane reguły. Tego typu działanie jest powszechne w przypadku człowieka i – jak dowodzi twierdzenie Gödla – niemożliwe w przypadku maszyny będącej przez swój formalny model w pełni zdeterminowaną. Jak zauważa Lucas, można w funkcjono-

---

[19] M. Hetmański, op. cit., s. 157.

[20] J. Lucas, *Minds, machines and Gödel*, „Philosophy” 1961, t. XXXVI, nr 137, s. 112–127.

waniu człowieka zauważyć również częste niekonsekwencje, nie dopatruje się on w tym jednak sprzeczności, lecz raczej wskazuje na jego omylność. Podkreśla za to pewną cechę: właściwe dla człowieka jest naprawianie swoich wcześniejszych błędów, refleksja nad zasadnością działania czy wreszcie weryfikacja prawdziwości jego twierdzeń. Tę specyficzną właściwość ludzkiego umysłu do wykraczania poza ramy własnej wiedzy, do ciągłego rozwoju i analizy własnych zachowań uznaje za cechę gatunkową, niemożliwą do odtworzenia przez jakiekolwiek maszyny.

> Znaczy to, że świadoma istota ludzka może zajmować się Gödlowskimi kwestiami w sposób, w który nie może maszyna, ponieważ świadoma istota może zarówno rozważać samą siebie, jak i swoje działanie i to nie inaczej niż w tym działaniu. Można powiedzieć o zbudowanej maszynie, że „rozważa" swoje własne działanie, lecz nie może ona wziąć go „pod rozwagę" bez stawania się tym samym inną maszyną, mianowicie starą maszyną z dodaną „nową częścią". Lecz to właśnie jest nieodłączne od naszej idei świadomego umysłu – tego, że może on odzwierciedlać siebie samego i krytykować swoje własne działania i do czego nie potrzebuje dodatkowej części; jest kompletny i nie ma żadnej pięty Achillesowej[21].

W 1965 r. w pracy *Alchemy and AI*[22] oraz w jej rozwinięciu z 1972 r. pod tytułem *What computers can't do*[23] Hubert Dreyfus poddał obliczeniową teorię umysłu silnej krytyce. W czasie pisania pierwszej pracował jeszcze w MIT, więc badaniom dotyczącym sztucznej inteligencji mógł się doskonale przyjrzeć. Swoją krytykę oparł na kilku filarach, odnosząc się do jego zdaniem głównych błędów i nieścisłości teorii AI. Jego opinie o komputacjonizmie, prócz zdroworozsądkowego charakteru, cechował również bardzo uszczypliwy, ironiczny język, co wywołało niemałe oburzenie w szeregach przedstawicieli kognitywnego nurtu, zaś wśród ich adwersarzy cieszyło się ogromnym powodzeniem. Argumenty Dreyfusa pogrupować można przez odniesienie ich do fundamentalnych założeń teorii AI.

Założenie biologiczne. Mózg to przetwarzający informacje mechanizm; charakter jego działania można przedstawić zasadą: włączony – wyłączony. Walter Pitts i Warren McCulloch, badając sieci neuronowe, doszli do wniosku, że działają one dwubiegunowo, stąd symulowanie ich pracy przez układ elektroniczny nie będzie stanowiło problemu. Tezę tę podważył Dreyfus, powołując się na badania w ramach neurologii przedstawiające analogowy (wielowartościowy) charakter sieci neuronowych, w przeciwieństwie do dwuwartościowego (0, 1) sygnału cyfrowego[24].

Założenie psychologiczne. Umysł można postrzegać jako mechanizm operujący cząstkami informacji, bazując na regułach formalnych. Przesłankę tę odrzuca

[21] Ibidem, s. 125.
[22] H. Dreyfus, *Alchemy and AI*, RAND Corporation 1965.
[23] H. Dreyfus, *What computers can't do*, MIT Press, New York 1972.
[24] Ibidem, s. 71–75.

Dreyfus, argumentując, że w dużej mierze na naszą wiedzę o świecie składa się szereg nastawień i tendencji, które kierują nas w stronę jednej interpretacji, w opozycji do innych. Nawet gdy używamy konkretnych symboli, kontrastujemy je z naszą zdroworozsądkową wiedzą, gdyż jedynie na jej tle symbole zyskują jakiekolwiek znaczenie. Nie została też ona zaimplementowana w jednostkowym umyśle pod postacią jednostkowych symboli o konkretnym znaczeniu[25].

Założenie epistemologiczne. Każda wiedza może być sformalizowana. Jak twierdził jeden z prekursorów AI, John Mccarthy, operująca symbolami maszyna może reprezentować każdą wiedzę, niezależnie od tego, czy człowiek reprezentuje ją tak samo. Zdaniem Dreyfusa, bezzasadność tego twierdzenia wynika bezpośrednio z odrzucenia założenia psychologicznego, ponieważ fakt występowania niesymbolicznej wiedzy (jak np. przeświadczenia) uniemożliwia formalne jej reprezentowanie[26].

Założenie ontologiczne. Na świat składają się niezależne fakty, które mogą być reprezentowane przez niezależne symbole. Zwolennicy teorii AI zakładają, że nie istnieje granica dla formalnego opisu wszechświata i każdy z jego fenomenów da się przedstawić i opisać za pomocą symboli i teorii naukowych. Konsekwencją tego założenia jest pogląd, że każde istnienie da się przedstawić jako obiekt, własność obiektu, jego klasę czy relację. Co więcej, da się je zrozumieć i opisać za pomocą logiki, matematyki i języka. W tym miejscu stawia Dreyfus jedynie pytanie: Jeśli założenie to jest nieprawdziwe, to gdzie leżą granice naszego poznania, możliwości inteligentnych maszyn i jaką rolę odegrać mogą one w naszym życiu?[27]

W pracy *Mind over machine*[28] Dreyfus analizuje zagadnienie podejmowania przez człowieka decyzji i możliwości symulacji tego procesu przez cyfrowe komputery, przy czym wyszczególnia dwie metody rozwiązywania problemów. Pierwszą z nich nazwał *knowing-that* (wiedzieć-że). Odnosi się ona do działania krok po kroku, a korzysta z niej człowiek, gdy na swojej drodze napotyka trudne do rozwikłania zagadnienie, które wymaga zatrzymania się i rozważenia konkretnych koncepcji jedna po drugiej. Jak przyznaje Dreyfus, nasze myślenie sprowadza się wtedy do manipulowania symbolami przy udziale logiki i języka. Ten rodzaj rozwiązywania problemów mogą zatem opanować cyfrowe maszyny, jednak z tej możliwości człowiek korzysta niezwykle rzadko.

Druga metoda rozwiązywania problemów jest oparta na właściwej człowiekowi kompetencji zwanej *knowing-how* (wiedzieć-jak). Jest to typowe, codzienne działanie niezwiązane z operowaniem jakimikolwiek symbolami, jak np. roz-

[25] Ibidem, s. 75–100.

[26] Ibidem, s. 101–117.

[27] Ibidem, s. 118–139.

[28] H. Dreyfus, *Mind over machine: The power of human intuition and expertise in the era of the computer*, Blackwell, Oxford 1986.

poznawanie twarzy, jazda samochodem czy znajdowanie odpowiednich słów w dyskusji. W takich sytuacjach człowiek od razu sięga po właściwą odpowiedź, bez wstępnego rozpatrywania wszystkich alternatyw. Działa więc intuicyjnie, często wręcz zapomina o wszystkich rządzących sytuacją regułach i spontanicznie wie, jak zareagować. W opinii Dreyfusa rozumienie sytuacji jest oparte na ludzkich celach, biologicznych ciałach i kulturze, warunkowanych przez nieświadome intuicje, przeświadczenia i wiedzę o świecie. Kontekst ten nie jest przechowywany w mózgu za pomocą symboli, tylko pod postacią szeregu intuicji. Wpływa również bezpośrednio na to, co zauważamy bądź nie, co uważamy za istotne, a co zupełnie pomijamy w naszej ocenie sytuacji. Tego rodzaju kontekstu, zdaniem Dreyfusa, nie może posiąść żadna skonstruowana przez człowieka maszyna cyfrowa. Nie może ona również rozwiązywać problemów w ten szybki intuicyjny sposób, gdyż z definicji skazana jest na niepowodzenie przez swoje formalne ograniczenia.

Przedłożone powyżej argumenty zdradzają sceptyczne nastawienie Dreyfusa względem materializmu w wersji redukcyjnej. Niejednokrotnie podkreślał on, że redukcja ludzkich stanów mentalnych do zachodzących w mózgu operacji jest z gruntu niemożliwa. Zwolennicy obliczeniowej teorii częstokroć podkreślają, że wraz ze stopniem złożoności i możliwości obliczeniowych komputerów wzrośnie także ich stopień podobieństwa do umysłu pod względem funkcjonalnym. Jak zauważa Dreyfus, zaobserwować można sytuację wprost przeciwną. Mianowicie od lat 70. (gdy po raz pierwszy wysunął swe wątpliwości w kierunku AI) na naszych oczach dokonuje się niezwykły postęp technologiczny w dziedzinie informatyki i elektroniki, co nie pociąga za sobą rozwiązania stawianych przed badaczami AI problemów. Zwielokrotnienie mocy obliczeniowej nie przyniosło przełomu w pracach nad stworzeniem sztucznej inteligencji, uwydatniając jedynie jej niedoskonałości. W ludzkim mózgu znajduje się $1{,}5–1{,}6\times10^{11}$ neuronów[29] i 10^{14} synaps[30], jednak ta ogromna złożoność nie sprawia, że działamy niezwykle szybko i bezbłędnie. Jak podkreśla Dreyfus, dzisiejsza nauka pozwala nam dostrzec, jak mozolnie pracuje nasz mózg w porównaniu do niektórych niezwykle zaawansowanych komputerów. W związku z tym to nie w złożoności organizacji systemu musi leżeć sekret fenomenu ludzkiego umysłu i ludzkiej świadomości. Intuicyjne działanie człowieka – jak powie Dreyfus – ma niewiele wspólnego z jakąkolwiek formą obliczania. Gdy jesteśmy pochłonięci jakimś działaniem niemal całkowicie „znika" nasza samoświadomość[31]. Ścigając uciekającą nam ulicą taksówkę, nie analizujemy stanów naszej świadomości. Jest tylko taksówka do złapania. Dreyfus, będąc prominentnym znawcą myśli Martina

[29] A. Longstaff, *Neurobiologia*, PWN, Warszawa 2012, s. 1–26.

[30] R.W. Williams, K. Herrup. *The control of neuron number*, „Annual Review of Neuroscience" 1988, nr 11, s. 423–453.

[31] H. Dreyfus, *What computers can't do*, s. 173.

Heideggera, podkreśla konieczność bycia w świecie jako podstawowy warunek dla powstania umysłu. Co więcej – powołując się na Heideggera – postuluje, że zdroworozsądkowej wiedzy nie przechowujemy w naszych mózgach, gdyż jest ona częścią świata, w którym żyjemy. Problem wiedzy, z jakim muszą się zmierzyć zwolennicy AI, Dreyfus nazywa mianem *commonsense knowledge problem*[32]. To swego rodzaju „ucieleśnienie" i jednocześnie „uspołecznienie" czy wręcz „uświatowienie" ludzkiego umysłu musi być wzięte pod uwagę, jeśli jego zagadkę mamy kiedyś rozwikłać. Na pytanie o naturę umysłu nie podaje jednoznacznej odpowiedzi, jednak z jego myśli wyłania się częściowa odpowiedź na postawione w niniejszym artykule pytanie.

Kluczowym zagadnieniem, które w tej formie znamy z kontrowersyjnych w środowisku filozoficznym prac Thomasa Nagela, jest to, w jaki sposób trzecioosobowa materialna konstrukcja (w mózgu bądź komputerze) może wykształcić pierwszoosobowy fenomen, by patrzeć na świat z pewnej szczególnej perspektywy[33]. Słynny eksperyment myślowy Nagela „jak to jest być nietoperzem" stawia zatem zdaniem Dreyfusa słuszne pytania, a obrany kierunek może doprowadzić do sensownych rozstrzygnięć. Kognitywiści często podkreślają, że przecież mózg w jakiś sposób to osiągnął. Jednak Dreyfus sądzi, że nie wytłumaczymy tego zjawiska, odwołując się do teorii masy krytycznej, zakładającej, że mózg osiągnął w pewnym momencie taką złożoność, że wygenerował umysł i świadomość. Dreyfus zgadza się jednak, że ewolucyjny rozwój naszego mózgu musi mieć wiele wspólnego z pojawieniem się fenomenu umysłu. Wysunięta przez niego krytyka komputacjonizmu z pewnością odrzuca wersję redukcyjną materializmu, jednak Dreyfus nie odrzuca pewnego powiązania pomiędzy biologicznym mózgiem a umysłem. W jego filozofii odnaleźć można różną niż u zwolenników AI wizję ludzkiego umysłu jako „czegoś więcej", niż zwykli oni mu przypisywać. Umysł jawi się więc w tym kontekście jako struktura złożona z wielu komponentów, zarówno tego materialnego, biologicznego, jak i intuicyjnego, społecznego oraz wielu jeszcze elementów, których dotąd nie wyjaśniliśmy.

Przyglądając się konsekwencjom krytyki obliczeniowej teorii, można zauważyć pewne punkty sporne w koncepcjach omawianych adwersarzy tego nurtu. Na pytanie o charakter zjawisk mentalnych udzielają zgoła różnych odpowiedzi, jednak z drugiej strony nie sposób nie zauważyć zasadniczych podobieństw. Wydaje się, że nie będzie zbyt pochopną generalizacją stwierdzenie, że w ramach kognitywistycznego nurtu zwykło się ostatnimi czasy spoglądać na fenomen umysłu jako pewną cechę, własność czy mechanizm ludzkiego mózgu (Dennett, Searle, Davidson). Wśród sekundantów obliczeniowej teorii to mózg uzyskuje pozycję nadrzędną nad umysłem jako jego pewną funkcjonalną strukturą. Adwersarze zaś

[32] H. Dreyfus, *Mind over machine*, s. 78.
[33] H. Dreyfus, *What computers can't do*, s. 182.

tych koncepcji ujmują umysł zgoła odmiennie. Wysuwając kolejne argumenty, często akcentują nadrzędną rolę tej rozwiniętej struktury nad zjawiskami zachodzącymi jedynie w naszym mózgu. Umysł jako niezwykle tajemnicza i nieuchwytna kompozycja wydaje się im czymś znacznie przekraczającym zmiany zachodzące w naszych sieciach neuronalnych. Próby przełamania kartezjańskiego dualizmu są w przytłaczającej większości podejmowane przez materialistów. Krytycy komputacjonizmu niemal jednym tchem odrzucają materializm redukcyjny, mniej lub bardziej sceptycznie spoglądając w stronę umiarkowanych ujęć materializmu nieredukcyjnego, czy to pod postacią emergentyzmu, czy rozmaitych wersji superweniencji. Dlatego należy zachowywać szczególną ostrożność w formułowaniu jednoznacznych rozstrzygnięć na temat ewentualnego sceptycyzmu wobec materialnej natury ludzkiego umysłu, oczywiście przy założeniu, że obliczeniową teorię odrzucimy.