Tytuł: **Genealogia i ideologia Sztucznej Inteligencji**

Autor: Grzegorz Borawski

Recenzent: dr hab. Andrzej Chmielecki

Źródło: http://www.kognitywistyka.net / mjkasperski@kognitywistyka.net

Data publikacji: 12 XII 2005

Pragnę złożyć serdeczne podziękowania dr hab. Andrzejowi Chmieleckiemu, którego wykłady zainteresowały mnie problematyką filozofii umysłu, za opiekę merytoryczną i użyczenie niezbędnej literatury oraz za wiele cennych uwag, dzięki którym niniejsza praca zyskała swój ostateczny kształt.

Wstęp

W latach pięćdziesiątych XX wieku, w wyniku rozwoju techniki komputerowej oraz pojawienia się nowych idei w językoznawstwie (Chomsky), ukształtowała się dziedzina badań oparta na idei tak zwanych „myślących maszyn". Dziedzinę tę określa się mianem *Sztucznej Inteligencji*.

Można rozróżnić dwa wymiary Sztucznej Inteligencji: operacyjny i ideologiczny. Wymiar operacyjny odnosi się do działań mających na celu praktyczne i teoretyczne korzyści, wynikające z prowadzonych badań. W tym aspekcie Sztuczna Inteligencja odnosi sukcesy i może znacznie poszerzyć zakres ludzkich możliwości poznawczych, przyczyniając się tym do rozwiązania wielu trudnych problemów (chociażby tych, które wiążą się z wykonywaniem skomplikowanych operacji medycznych).

Badania nad *sztuczną inteligencją*[1] mogą budzić wątpliwości tylko w przypadku, gdy nadaje się im cel ideologiczny, gdy cele poznawcze i praktyczne ustępują miejsca pragnieniu stworzenia artefaktu, który pewnymi sprawnościami mógłby nie tylko przypominać żywe stworzenia, ale nawet przewyższyć człowieka pod względem zdolności intelektualnych.

W niniejszej pracy nie będę poruszał kwestii związanych z operacyjnym, czy technicznym wymiarem badań nad sztuczną inteligencją, lecz skupię się na najważniejszych ideologicznych trudnościach Sztucznej Inteligencji. Wynikają one w dużej mierze z

[1] Wprowadzam tu rozróżnienie zaproponowane przez M.J. Kasperskiego: „termin pisany wielką literą (Sztuczna Inteligencja) określać będzie dziedzinę badań (…), a oddany małymi literami (…) – przedmiot jej badań" – M.J. Kasperski, *Sztuczna Inteligencja*, Helion, Gliwice 2003, s. 20.

unaukowienia koncepcji umysłu, a wiążą się z takimi zagadnieniami, jak problem obliczalności zjawisk mentalnych, czy jakościowy aspekt ich treści.

Według mnie, punktem wspólnym tych trudności jest błędne pojęcie rozumienia, zbyt szerokie pojęcie umysłu (jako wszystkiego, co nie-fizyczne) przy jednoczesnym ograniczeniu pojęcia myślenia do czynności umysłowych związanych z przetwarzaniem symboli. Nie można również pominąć wagi teoretycznego spadku po tych koncepcjach psychologicznych, które same borykały się z wieloma problemami.

Są to zagadnienia wchodzące w zakres filozofii umysłu, podejmowane w ostatnich kilkudziesięciu latach przez wielu filozofów, do których odnoszę się w tej pracy. Ukazuję w niej najważniejsze koncepcje, które przyczyniły się do ukształtowania ideologicznych podstaw Sztucznej Inteligencji. Ideologię pojmuję tu jako pozamerytorycznie uwarunkowany zespół przekonań opisujących rzeczywistość, właściwy danej grupie ludzi i kierunkowi prowadzonych przez nich badań. Przekonania te utrzymywane są zwykle z powodów innych, niż racje poznawcze. Zwolennicy ideologii mogą kierować się na przykład własnym interesem i irracjonalnymi pragnieniami realizowania swych marzeń; mogą także opierać się na sile autorytetów, nie przyjmując do wiadomości, że nawet wybitny filozof może się mylić. Dlatego też ideologia, tworząca pewien obraz świata, prezentuje również wizję przyszłości, która jest w mniejszym lub większym stopniu „upiększona”. Wizja ta pozwala wyznawcom ideologii wyznaczać cele, które mają uzasadniać określone działania. I z tym właśnie mamy do czynienia w przypadku, gdy pominiemy praktyczne przesłanki prowadzenia badań nad sztuczną inteligencją.

Czy jest możliwa realizacja „upiększonej” wizji przyszłości, w której myślące komputery będą we wszystkim wyręczać człowieka? A może myślące maszyny będą stanowić dla ludzi zagrożenie? Czy komputery w ogóle mogą myśleć? Będę dążył do udzielenia odpowiedzi na ostatnie z wymienionych pytań. Wymaga to sięgnięcia do korzeni, na których wyrosła ideologia Sztucznej Inteligencji, czemu poświęcona jest cała pierwsza część pracy.

Centralnym problemem w filozofii umysłu jest wyjaśnienie relacji zachodzących między tym, co umysłowe a tym, co fizyczne. Aby ukazać, jak doszło do sformułowania tego problemu, przedstawiam starożytną koncepcję relacji duszy i ciała, dokonując tego na przykładzie dwóch wybitnych filozofów: Platona i Arystotelesa. Wybór ten nie jest przypadkowy. Platon wyraził bowiem myśl ważną w polemice z ideologią Sztucznej Inteligencji, natomiast Arystoteles bywa uznawany za prekursora funkcjonalizmu, stanowiącego swego rodzaju szkielet tej ideologii. Sformułowaną przez Stagirytę koncepcję duszy rozumnej można uznać za podstawę przedstawionej w ostatnim rozdziale definicji myślenia, a ponadto, bez tej koncepcji, nie narodziłby się prawdopodobnie kartezjański racjonalizm.

Nowożytny racjonalizm dał początek rozważaniom, których nie sposób pominąć, jeśli chce się poznać korzenie filozofii umysłu. Dualistyczna wizja istoty ludzkiej, jaką zaprezentował Kartezjusz, okazała się inspirująca dla wielu filozofów, którzy zmagali się z pytaniami dotyczącymi relacji zachodzących między substancją duchową a cielesną.

Jeśli kartezjański interakcjonizm uzna się za teorię wątpliwą (a tak właśnie czyniono), to pozostają jeszcze trzy możliwości ustalenia tych relacji. Można mianowicie przyjąć, że byty materialne i duchowe współdziałają, wzajemnie na siebie nie oddziałując, czego przykładem jest paralelizm psychofizyczny Malebranche’a. Można także stwierdzić, że dusza ma wpływ na ciało, ale substancja materialna nie ma zdolności oddziaływania na to, co duchowe.

Stanowisko takie określa się mianem animizmu, zaś jego przeciwieństwem jest epifenomenalizm.

Innym podejściem do problemu psychofizycznego jest zaprzeczenie temu, że dusza i ciało istnieją jako odrębne substancje. Uczynił tak Spinoza, uznając myślenie i rozciągłość za dwa z wielu atrybutów jedynej substancji doskonałej, boskiej. Można jednak także zaprzeczyć istnieniu substancji duchowej, czego przykładem jest monizm materialistyczny Hobbesa.

Materializm, ukazany w świetle koncepcji przedstawionej przez Spinozę, przyczynił się do nowego sformułowania problemu psychofizycznego. Współcześnie nie bierze się już pod uwagę istnienia substancji duchowej, a stawia się pytania dotyczące relacji zachodzących pomiędzy *własnościami* umysłowymi a własnościami fizycznymi. W miejsce dualizmu substancjalnego pojawił się dualizm własności.

W pracy tej przedstawiam również stanowisko przeciwne materializmowi, co uważam za niezbędne dla ukazania genealogicznych i ideologicznych podstaw Sztucznej Inteligencji. Chodzi mianowicie o monadologię Leibniza, a przede wszystkim o jego koncepcję „maszyny myślącej", funkcjonującej na podstawie uniwersalnego systemu językowego, który w pełni podlega zasadom arytmetyki. Tkwiąca w pomysłach Leibniza idea gramatyki uniwersalnej i mechanicystyczna wizja człowieka, jaką zaprezentował między innymi La Mettrie, stały się inspiracją dla teoretyków Sztucznej Inteligencji.

Duże znaczenie w tej kwestii należy również przypisać reprezentacjonizmowi Locke'a oraz asocjacjonizmowi Hume'a. Koncepcje te przyczyniły się do powstania kognitywistyki oraz funkcjonalistycznej koncepcji umysłu. Warto tu jednak zaznaczyć, że nie chodzi o nurt zwany *funkcjonalizmem*, który pojawił się w psychologii i kojarzony jest zwykle z jego twórcą, Johnem Dewey'em (1859–1952) albo psychologiem i filozofem, Williamem Jamesem (1842–1910). Funkcjonalizm psychologiczny w tym ujęciu kładł nacisk na problem funkcjonowania organizmu w środowisku i interakcji człowieka z jego otoczeniem. W niniejszej pracy zajmuję się funkcjonalizmem, który narodził się w połowie XX wieku, jako koncepcja filozoficzna, nawiązująca między innymi do behawioryzmu, któremu poświęcam osobny rozdział. Dopiero na bazie tego stanowiska oraz nadmienionego funkcjonalistycznego podejścia w psychologii, narodził się funkcjonalizm, traktowany jako odrębna teoria naukowa (psychofunkcjonalizm).

Nie chodzi zatem w tej pracy o funkcjonalistyczny nurt w psychologii, lecz o funkcjonalizm w filozofii umysłu i funkcjonalizm ujęty jako teoria psychologiczna, a przede wszystkim, jako ideologiczna podstawa dla prowadzenia badań nad sztuczną inteligencją, czerpiąca w dużej mierze z komputacyjnych koncepcji Alana Turinga i lingwistyki Noama Chomsky'ego. Nie bez powodu więc spora część pracy zawiera treści dotyczące różnych odmian funkcjonalizmu i problemów, które ta teoria wnosi do ideologii Sztucznej Inteligencji.

W funkcjonalizmie uwidacznia się podstawowy podział, jakiemu podlega fizykalizm, czyli współczesna materialistyczna wizja świata. Mechanicyzm i tak zwaną *teorię identyczności* należy zaliczyć do redukcjonistycznej odmiany materializmu, utożsamiającej własności umysłowe z fizycznymi. Alternatywną odmianą materializmu jest fizykalizm nieredukcjonistyczny, dla którego dualizm własności stanowi taki sam problem, jak dualizm substancjalny dla filozofów nowożytnych. Próbuje się go rozwiązać na różne sposoby. Postuluje się między innymi, że własności umysłowe są nadbudowane na fizycznych i w ten sposób od nich zależne. Jest to zgodne z teorią superweniencji, której konsekwencją jest

jednak epifenomenalizm własności mentalnych. Aby takich konsekwencji nie ponosić, można za emergentystami przyjąć, że własności umysłowe, jako „wyłaniające się" z własności fizycznych, oddziałują przyczynowo na to, co fizyczne. Można również uznać, że zdarzenia mentalne superweniują na fizycznych, ale tylko w ten sposób, że stanowią ich nie-naukowy opis. Jest to twierdzenie zgodne z zasadami monizmu anomalnego. Mamy wtedy do czynienia ze swego rodzaju „epifenomenalizmem pojęciowym". Wymienione teorie stanowią uzupełnienie funkcjonalistycznej koncepcji umysłu, leżąc tym samym u podłoża ideologii Sztucznej Inteligencji.

W drugiej części pracy odnoszę się już bezpośrednio do problematyki związanej z „myślącymi maszynami". Za przedmiot rozważań obieram przy tym wyłącznie maszyny cyfrowe. Najpierw, przedstawiając podstawowe argumenty przeciw Sztucznej Inteligencji, nie wykraczam poza obręb rozważań dających się ująć jako „semiotyczne". Następnie, w trzeciej części, posiłkując się koncepcją umysłu opartą na pewnej wersji teorii informacji i ukazując zasadnicze różnice między komputerem a ludzkim umysłem, odpowiadam na najbardziej drażliwe dla ideologów Sztucznej Inteligencji pytanie: *czy maszyny mogą myśleć?*

Część I
Genealogia Sztucznej Inteligencji

Rozdział I
Od duszy do obliczania

1.1. Relacja duszy i ciała w filozofii starożytnej

1.1.1. Platon

Słynne jest platońskie ujęcie ciała jako więzienia dla duszy[2]. Taka koncepcja nakazuje uważać duszę za byt, który może się oddzielić od ciała. Jak twierdzi Giovanni Reale, Platon nigdy nie zarzucił dualistycznej koncepcji relacji zachodzących między duszą a ciałem, co można wyjaśnić tym, że do koncepcji tej „oprócz składnika metafizyczno-ontologicznego dochodzi religijny składnik orfizmu, który strukturalne rozróżnienie między duszą (=nadzmysłową) i ciałem (=zmysłowym) przekształca w *strukturalną opozycję*"[3].
Tradycyjna wykładnia platonizmu opiera się na dosłownej, mitologicznej interpretacji nauk „pisanych" Platona. Według tej wykładni, obiektywnie i odrębnie istnieją świat idei i wtórny wobec niego świat rzeczy zmysłowych; idee stanowią wzorce dla rzeczy zmysłowych; świat

[2] „Ci, którzy kochają naukę, poznają, że filozofia duszę ich znajduje po prostu związaną i przyrosłą do ciała, i przymuszoną oglądać byty przez ciało niby przez kraty więzienia: dusza ich nie może patrzeć sama przez się i w tej ciemnocie swojej wije się i widzi, jak straszne jest to więzienie, w którym ją żądze trzymają, a sam więzień pomaga własne zacieśniać kajdany – (…) błędu pełne jest poznanie przez oczy i pełne błędu owo przez uszy i przez inne wrażenia zmysłowe, (…) a co by za czyimś pośrednictwem dostrzegła, tu takie, a tam inne, tego niech za żadną prawdę nie uważa. Takim jest wszystko to, co zmysłami dostrzegalne i widzialne, a to, co ona sama widzi, to umysłem tylko pojęte i pozbawione postaci" – Platon, *Fedon*, 82 E, w: Platon, *Dialogi*, t. I, tłum. W. Witwicki, Wyd. Antyk, Kęty 1999, s. 666.
[3] G. Reale, *Historia Filozofii Starożytnej*, tom II, RW KUL, Lublin 1996, s. 242.

idei ma strukturę hierarchiczną – według poziomu ogólności idei, gdzie najogólniejsza jest idea Dobra.

Jak zauważa Reale, mamy tu do czynienia z rozróżnieniem dwóch poziomów rzeczywistości: poziomu rzeczy widzialnych (postrzegalnych zmysłami) i poziomu rzeczy niewidzialnych (możliwych do uchwycenia tylko za pomocą intelektu).

Sama dusza była dla Platona *ideą*, a mianowicie „ideą życia", czyli tym, „co ze swej natury jest życiem i daje życie"[4], a więc nie może być zniszczalne i martwe – musi być wieczne. A skoro „mówić o życiu znaczy mówić o ruchu"[5], to pojęcie duszy jako zasady ruchu pozwala w pewien sposób dowieść jej nieśmiertelności[6]. Dowody te wskazują jednak raczej na wieczność idei, a nie duszy pojmowanej jako „człowiek bez ciała", której nieśmiertelność może być ewentualnie kwestią wierzeń (choćby orfickich).

Szczególnie ważnym dla tematu tej pracy jest ukazanie przez Platona idei jako formalnych wzorców, a nie przyczyn materialnych, czy sprawczych. W koncepcji tej istotne jest rozróżnienie materialnych *czynników realizacji* i formalnych *czynników determinacji*, ukazane wyraźnie w następującej wypowiedzi Sokratesa: „Jeśliby ktoś powiedział, że nie mając tego wszystkiego, tych kości i ścięgien i co tam innego mam jeszcze, nie byłbym w stanie zrobić tego, co mi się podoba, miałby słuszność. Ale powiedzieć, że ja dlatego i przez to robię to, co robię, i przez to się rozumem kieruję, a nie przez to, że wybieram to, co najlepsze, to byłaby wielka i niebywała lekkomyślność. To by znaczyło nie umieć rozróżnić, że czymś innym jest przyczyna czegoś, co istnieje, a czymś innym to, bez czego przyczyna nie byłaby przyczyną. Mam wrażenie, że to właśnie wielu ludzi, jakby po ciemku macając, nazywa przyczyną samą, choć ta nazwa czemuś innemu przynależy"[7].

Chciałoby się rzec, że ci „po ciemku macający" ludzie to zwolennicy poglądów materialistycznych, uznający ciało (a właściwie mózg) za jedyną przyczynę „kierowania się rozumem" (myślenia). W powyższym cytacie Platon sprzeciwia się takiemu podejściu i kładzie nacisk na rolę form, jako czynników determinacji, które należy odróżnić od czynników realizacji, stanowiących przyczyny materialne i sprawcze. Według Platona, ludzkie działania są *zdeterminowane* (określone) przez rozum (człowiek kieruje się w swych działaniach rozumem). Ciało nie jest przyczyną ludzkich poczynań, chociaż bez formalnie określonej materii nie mogłyby one zostać zrealizowane i rozum nie mógłby ich determinować.

Ukazana przez Platona rola formy, jako czynnika determinacji, zostanie później uwypuklona w semiotycznych rozważaniach na temat myślenia. Aby ukazać, w jaki sposób forma jakiejś całości determinuje części tej całości, można przedstawić przykład następującego zdania, opisującego działanie językoznawcy, któremu „udał się podział systemu językowego, ale gdzieś podział się sens, więc udał się na jego poszukiwania". To właśnie syntaktyczna struktura zdania, czyli jego forma, określa znaczenie użytych w tym zdaniu słów *udał* i *podział*.

[4] *Ibidem*, s. 225.
[5] *Ibidem*, s. 226.
[6] „Wszelka dusza jest nieśmiertelna. Bo co się wiecznie rusza, nie umiera" – Platon, *Fajdros*, 245c-246a, w: Platon, *Dialogi*, t. II, tłum. W. Witwicki, Wyd. Antyk, Kęty 1999, s. 138.
[7] Platon, *Fedon*, 99 B, w: Platon, *Dialogi*, t. I, s. 688.

Czy jednak uprawnione jest należące do ideologii Sztucznej Inteligencji twierdzenie, że syntaktyka określa semantykę? Odpowiedzi na to pytanie udzielę w trzecim rozdziale tej pracy. Tutaj pozwolę sobie jedynie przytoczyć zdanie zaczerpnięte z wykładów Andrzeja Chmieleckiego: „prawo determinuje fakty, choć może ono być poznane tylko poprzez znajomość faktów". Znajomość faktów można by analogicznie odnieść do znajomości wielu znaczeń, które mogą mieć wyrazy *udał* i *podział*, natomiast znajomość prawa może dotyczyć reguł syntaktycznych, według których powyższe zdanie zostało uformowane. Gdy stwierdzi się, że znajomość czynników determinacji wymaga znajomości czynników realizacji, zarysowuje się odpowiedź na postawione wyżej pytanie. Udzielenie jej nie byłoby możliwe bez platońskiego pojęcia przyczyny (*aitia*), jako czynnika determinacji, a nie realizacji.

1.1.2. Arystoteles

Giovanni Reale[8] twierdzi, że choć cenny u Platona jest fakt dostrzeżenia przez tego filozofa idealnej natury duszy, to rację należy również przyznać presokratykom, którzy uważali, że dusza nie może odłączyć się od ciała. Arystoteles przedstawił swoistą syntezę tych dwóch stanowisk, ukazując każdą żywą istotę jako jedność materii i formy. Twierdzenie to ugruntowane jest w koncepcji hylemorficznego złożenia wszystkich bytów z materii jako możności i formy jako aktu[9]. Dusza to według Stagiryty forma substancjalna ciała (*eidos*) czyniąca ciało żywym. Dzięki niej ciało jest tym, czym jest[10].

Arystoteles zwrócił uwagę, że w opisie ciała należy uwzględnić nie tylko jego podstawę materialną, ale również funkcje przez to ciało spełniane i z tego właśnie powodu bywa uważany za prekursora współczesnego funkcjonalizmu. Orientację tę zapoczątkował Hilary Putnam[11] stwierdzając, że „tym, co nas najbardziej interesuje, jak rzekł Arystoteles, jest forma, a nie materia"[12]. Powoływał się przy tym na cytat z dzieła *O duszy*, w którym Stagiryta uznaje, że „nie ma uzasadnienia pytanie: czy dusza i ciało stanowią coś jednego, jak [nie ma sensu pytać], czy wosk i odcisk na nim [stanowią coś jednego] i w ogóle materia jakiejkolwiek rzeczy i *to, czemu ona służy jako materia*"[13] [zaznaczenie kursywą – G.B.].

Bardziej jednak adekwatny dla traktowania Arystotelesa jako prekursora funkcjonalizmu wydaje się przykład przytoczony przez Józefa Bremera, w którym zauważa on, że według Stagiryty, oko pozbawione swej funkcji, czyli zdolności widzenia, w ogóle nie jest okiem. Można powiedzieć, że „zdolnością widzenia" ciała jest dusza, będąca dla Arystotelesa zasadą

[8] G. Reale, *op. cit.*, s. 454.

[9] Jak zauważa Andrzej Chmielecki, Arystoteles posługiwał się dwoma pojęciami formy: *eidos* i *morfe*. W pierwszym znaczeniu forma była rozumiana jako „ogół określeń istotnych (atrybutów) (…) bytu" i zarazem „czynny, kształtujący (…) pierwiastek bytu, który sprawia, że rzecz jest tym, czym jest", a więc stanowi o istocie tego bytu. Pod pojęciem formy jako *morfe* Arystoteles miał na myśli organizację, ukształtowanie, strukturę, „układ relacji wiążących ze sobą poszczególne części danego bytu" – A. Chmielecki, *Rzeczy i wartości*, PWN, Warszawa 1999, s. 62-63.

[10] Dusza jest według Arystotelesa „substancją w znaczeniu formy, która decyduje o istocie (…) ciała. Tak na przykład, gdyby jakieś narzędzie, powiedzmy siekiera, było ciałem naturalnym, odpowiedź na pytanie: «czym jest siekiera» stanowiłaby jej istotę, a w następstwie tego jej duszę" (Arystoteles, *O duszy*, III, 412b, w: Arystoteles, *Dzieła wszystkie*, tom III, tłum. P. Siwek, PWN, Warszawa 1992, s. 71). „Duszą" siekiery jest w takim pojęciu określona forma, która czyni ją „narzędziem zdolnym do cięcia drzewa" (*Ibidem*, przypis, s. 71).

[11] Hilary Putnam „wprowadził funkcjonalizm w późnych latach sześćdziesiątych" – J. Kim, *Umysł w świecie fizycznym*, tłum. R. Poczobut, PAN, Warszawa 2002, s. 112.

[12] H. Putnam, *Philosophy and our mental life*, w: H. Putnam, *Mind, Language and Reality*, Cambridge University Press, Cambridge 1975, s. 302.

[13] Arystoteles, *O duszy* II, 412b, w: Arystoteles, *op. cit.*, 1992, s. 71.

(*arche*) życia, „bez której ciało nie jest ciałem, tylko zlepkiem części materialnych”[14] i która, jako zasada ruchu, „*umożliwia* jego poruszanie się, *poprzedza* stany funkcjonalne i *warunkuje* ich realizację”[15] [zaznaczenie kursywą – G.B.]. Dusza nie jest w tym pojęciu funkcją ciała, lecz dynamiczną harmonią (*morfe*), dzięki której ciało zorganizowane jest tak, że może spełniać określone funkcje związane na przykład z rozumnym działaniem.

Istotą człowieka była więc u Arystotelesa dusza jako forma substancjalna, a nie funkcje ciała, co może podważać zasadność przypisywania Arystotelesowi miana pierwszego funkcjonalisty. Aby być funkcjonalistą, nie wystarczy mówić o funkcjach, lecz trzeba pójść dalej i uznać je za istotę czegoś. Dlatego samo funkcjonalne rozróżnienie poszczególnych władz duszy nie czyni Arystotelesa funkcjonalistą.

Arystoteles przedstawił trójpodział duszy oparty na obserwacji podstawowych zjawisk i funkcji życiowych. Wyróżnił duszę wegetatywną (odnoszącą się do takich zjawisk i funkcji, jak rodzenie, czy wzrost), duszę zmysłową (związaną z postrzeganiem i ruchem) i duszę rozumną (odpowiedzialną za namysł i poznanie). Pierwszą z nich mają rośliny, pierwszą i drugą – zwierzęta, a wszystkie trzy – człowiek. Należy przy tym pamiętać, że powyższy podział nie wyodrębnia trzech jakościowo różniących się od siebie dusz, lecz jest podziałem duszy według jej funkcji, wyróżnieniem jej władz. Można tu raczej mówić o podziale strukturalnym, w którym funkcja poszczególnych władz jest określana przez ich miejsce w strukturze duszy. Architektura duszy (*morfe*) determinuje więc jej funkcje. Kiedy mówi się o umyśle, w grę wchodzi oczywiście dusza rozumna.

Arystoteles ujmował poznanie intelektualne w kategoriach możności i aktu. Intelekt określony został jako zdolność i możność poznania czystych form zawartych w postrzeżeniach zmysłowych i wyobraźni. Można wobec tego stwierdzić, że za źródło poznania Arystoteles uznawał wrażenia zmysłowe[16]. Myślenie ukazane zostało nie tylko jako wynik powtarzających się i łączących ze sobą wyobrażeń, z czym mamy do czynienia już u zwierząt, lecz przede wszystkim jako wymagające zdolności „sprowadzania większej ilości wyobrażeń do jedności”[17].

Relacja ciała i duszy nie stanowiła u Arystotelesa problemu. Została ona ujęta teleologicznie – poprzez określenie duszy jako źródła ruchu umożliwiającego celowy rozwój organizmu. Dopiero mechanicystyczne rozumienie ciała ożywionego dało początek problemowi psychofizycznemu, zmuszając do poszukiwania źródła ruchu ciała poza nim samym. Jak zauważa Jerzy Bobryk, ruch był w koncepcji Arystotelesa „aktualizowaniem się wewnętrznej potencjalności. Dla Kartezjusza człowiek był duchem uwięzionym w maszynie”[18].

[14] J. Bremer, *Problem Umysł-Ciało*, WAM, Kraków 2001, s. 37.
[15] *Ibidem*, s. 39.
[16] Można w związku z tym uznać Arystotelesa za przedstawiciela empiryzmu genetycznego.
[17] Stwierdzenie to (zawarte w: Arystoteles, *O duszy*, III, 434a, w: Arystoteles, *op. cit.*, 1992, s. 142) będzie miało zasadnicze znaczenie dla podstawowego kontrargumentu wobec koncepcji myślących maszyn cyfrowych. Za jedyną władzę zdolną do uchwycenia formy w wyobrażeniach Arystoteles uznał władzę intelektualną, której posiadania nie można przypisywać komputerom, co uzasadnię w trzecim rozdziale pracy.
[18] J. Bobryk, *Locus umysłu*, PAN, Wrocław 1987, s. 15.

1.2. Relacja umysłu i ciała w filozofii nowożytnej

1.2.1. Interakcyjny dualizm kartezjański

Kartezjusz przyczynił się do narodzin mechanistycznej wizji świata, która zaczęła się kształtować dzięki naukom Kopernika i Galileusza[19]. To, że Rene Descartes (1596-1650) pojmował ciało mechanistycznie i skupił się w swych rozważaniach na „duszy rozumnej" (jeśli trzymać się starożytnych podziałów), było znakiem czasów, w których żył. Wiek XVII można porównać do wielkich wieków starożytności pod względem bogactwa treściowego i osobowości filozoficznych. Ówczesny rozwój nauk szczegółowych wpłynął na charakter zapoczątkowanej przez Kartezjusza filozofii nowożytnej. Można określić tego filozofa mianem „ojca racjonalizmu i interakcyjnego dualizmu psychofizycznego", czego najlepszym uzasadnieniem są treści zawarte w jego dziełach, spośród których na szczególną uwagę zasługuje nie tylko słynna *Rozprawa o metodzie* (1637), ale także *Medytacje o pierwszej filozofii* (1641), *Zasady filozofii* (1644) i spora część korespondencji prowadzonej z wiodącymi myślicielami tamtych czasów.

Największym i najtrudniejszym do zdobycia dobrem, dla którego warto poświęcić się bez reszty, była dla Kartezjusza *Prawda*, rozumiana przezeń jako to, co „rozum jasno i wyraźnie pojmuje". Uważał przy tym, że jedyną dyscypliną zasługującą na miano nauki jest matematyka, gdyż tylko ona dysponuje właściwą *metodą* badawczą. Określenie właściwej metody było dla Kartezjusza kluczem do zmiany źle przez niego ocenianej sytuacji w świecie nauki. W *Rozprawie o metodzie* Descartes przedstawił założenia metody, której zastosowanie miało gwarantować bezbłędność działań poznawczych[20].

Według Kartezjusza ludzki rozum posiada właściwość odróżniania prawdy od fałszu w sposób analogiczny do odróżniania przez organy zmysłowe kształtów i dźwięków. Zmysły jednak niejednokrotnie mogą wprowadzać w błąd, co wskazuje według Kartezjusza na niepewność zdań empirycznych. Uznał on, że aby uzyskać wiedzę pewną, należy poddać wiedzę dotychczasową radykalnej krytyce, gdyż jego zdaniem, tylko poprzez radykalne wątpienie można osiągnąć prawdę niewątpliwą.

Sceptycyzm kartezjański, jako instrument umożliwiający osiągnięcie prawdy niewątpliwej, był sceptycyzmem metodycznym. Poprzez uznanie wątpienia za szczególny przypadek myślenia, sceptycyzm ten doprowadził Kartezjusza do twierdzenia „*cogito ergo sum*", wskazującego na konieczność istnienia podmiotu myślącego, którego istotą nie jest cielesność, lecz myślenie. Z tego rozumowania wyłoniła się koncepcja podziału rzeczywistości na byty materialne i duchowe oraz, będąca tego konsekwencją, wizja

[19] Już tutaj warto zauważyć ogromny wpływ Kartezjusza na współczesną wizję świata: „W 1994 roku papież Jan Paweł II stwierdził, że to właśnie Kartezjusz, być może nieświadomie, zapoczątkował etap rozpadu średniowiecznej, chrześcijańskiej wizji świata i zastąpił ją aparaturą pojęciową, która przyspieszyła rozkwit racjonalizmu, a w konsekwencji – zepsucie powodowane przez nowoczesność i «śmierć Boga»" – cytat za: H.M. Bracken, *Kartezjusz* w: *Historia filozofii zachodniej* pod red. R.H. Popkina, tłum. A. Zbrzezny, Zysk i S-ka, Poznań 2003, s. 354. Stwierdzenie „być może nieświadomie" ma tu duże znaczenie, ponieważ Kartezjusz nie odrzucał prawd religijnych, lecz pragnął ująć je z punktu widzenia dyscypliny intelektualnej obowiązującej na terenie nauki.

[20] Uwidacznia się tu charakterystyczny dla ówczesnej postawy filozoficznej *maksymalizm*, który w swej matematycznej odmianie wynikł ze stwierdzenia przez Kartezjusza pozorności podziału nauk, do czego skłoniło go odkrycie braku różnicy przedmiotowej między algebrą a geometrią (Kartezjusz był twórcą geometrii analitycznej).

człowieka jako jedynego bytu złożonego z ciała jako substancji rozciągłej (*res extensa*), podlegającej prawom mechaniki, oraz z umysłu, czyli duszy, jako substancji myślącej (*res cogitans*), która jest niepodzielna i nie posiada atrybutu rozciągłości[21].

„Myślenie (*cogito*) staje się istotą tego co niecielesne, antytezą wszystkiego co cielesne[22]. Na miejsce Arystotelesowskiej duszy pojawia się świadomy siebie samego umysł"[23]. Tak więc, oprócz twierdzącej odpowiedzi na pytanie, czy dusza *istnieje*, Kartezjusz określił, *czym* dusza jest. Uczynił to wskazując, że istotą duszy jest myślenie[24]. Dusza przestała być starożytną „zasadą życia" i okazała się wyłącznie „duszą rozumną" (umysłem).

Metafizyczna różnica między umysłem a ciałem wskazuje, że każdą z tych substancji można poznać bez odnoszenia się do drugiej[25]. Descartes przypisywał materii wyłącznie właściwości geometryczne i mechaniczne, odwołując się w wyjaśnianiu zjawisk przyrodniczych tylko do matematycznie opisywanych związków przyczynowych, a eliminując w tym zakresie aspekt celowości[26]. Według Kartezjusza, kategoria przyczyny celowej nie może mieć zastosowania w odniesieniu do świata materialnego, znajdując zastosowanie wyłącznie do opisu świata duchowego. Ciało można poznawać za pomocą zmysłów zewnętrznych[27], a umysł poprzez wolne od zwątpienia poznanie wewnętrzne[28], przy czym źródłem pewności może być tylko rozum[29].

Kartezjusz określił kryteria, którym musi sprostać wiedza, aby mogła być uznana za wiedzę autentyczną: musi ona być niepodważalna (pewna) i ugruntowana w ideach[30]. Wynika z tego,

[21] „Tak, z przyczyny, iż zmysły nasze zawodzą nas niekiedy, przyjąłem, że żadna rzecz nie jest taka, jak one nam przedstawiają. Ponieważ istnieją ludzie, którzy mylą się w rozumowaniu (…), pomyślałem, iż ja jestem podległy błędom równie, jak każdy inny, i odrzuciłem, jako fałszywe, wszystkie racje, które przyjąłem niegdyś, jako dowiedzione. (…) Ale równocześnie zastanowiłem się, iż, podczas gdy silę się przypuścić, że wszystko jest fałszywe, trzeba, abym ja, który myślę, był czemś; i, zważając, iż ta prawda: MYŚLĘ, WIĘC JESTEM, jest (…) pewna i niezłomna (…), osądziłem, iż mogę ją przyjąć (…) za pierwszą zasadę filozofii. (…) mogę udać, jakobym nie miał ciała i jakoby nie było żadnego świata ani miejsca, gdzieżbym był; nie mogę wszelako udać, jakobym nie istniał. (…) Poznałem stąd, że jestem substancją, której całą istotą lub przyrodą jest jeno myślenie, i która, aby istnieć, nie potrzebuje żadnego miejsca, ani nie zależy od żadnej rzeczy materjalnej; tak, iż to JA, to znaczy dusza, przez którą jestem tem, czem jestem, jest zupełnie odrębna od ciała, a nawet jest łatwiejsza do poznania, niż ono, i że gdyby nawet ono nie istniało, byłaby i tak wszystkiem, czem jest." – Kartezjusz, *Rozprawa o metodzie*, tłum. T. Boy-Żeleński, De Agostini, Warszawa 2002, s. 90-91.

[22] Według Johna Searle'a podział ten „pokutuje" do dziś w postaci „dualizmu pojęciowego", sprowadzającego się do „stwierdzenia, że w pewnym ważnym sensie «fizyczne» implikuje «niementalne», a «mentalne» implikuje «niefizyczne». Zarówno tradycyjny dualizm, jak tradycyjny materializm zakładają tak pojęty dualizm pojęciowy." – J. R. Searle, *Umysł na nowo odkryty*, PIW, Warszawa 1999, s. 47.

[23] J. Bremer, *op. cit.*, s. 45.

[24] Kartezjańska substancja duchowa jawi się bardzo tajemniczo. Oprócz tego, że jej istotą jest myślenie i że mieści się w szyszynce, można ją opisywać tylko poprzez stwierdzanie, czym *nie jest*.

[25] Arystoteles również metafizycznie rozróżniał ciało i duszę (materię i formę), ale nie uznał ich za dwie odrębne substancje.

[26] Tak jak Demokryt, a w przeciwieństwie do Arystotelesa, który przypisywał materii zdolność do celowego rozwoju.

[27] Poprzez takie *modi*, jak spoczynek – ruch, postać (za: J. Bremer, *op. cit.*, s. 44).

[28] Poprzez takie *modi*, jak wiedzieć – nie wiedzieć, wątpić – nie wątpić, chcieć – nie chcieć, wyobrażać sobie (za: *ibidem*).

[29] Ze względu na postawę radykalnego racjonalisty, Kartezjusz został uznany za wroga Kościoła katolickiego, chociaż prawomocność jego dowodu istnienia Boga (opartego nie na objawieniu, lecz na kategorii związku przyczynowo-skutkowego) dorównywała prawomocności dowodów św. Tomasza z Akwinu.

[30] Koncepcję *idei* Kartezjusz zaczerpnął od Platona, który posiadanie wiedzy apriorycznej tłumaczył anamnezą: „jeśli więc, uważam, dostawszy ją przed urodzeniem, zapomnieliśmy ją przychodząc na świat, a potem, posługując się zmysłami, na powrót tamte wiadomości odzyskujemy, któreśmy przedtem kiedyś posiadali, to czyż to, co nazywamy uczeniem się, nie jest odzyskiwaniem naszej własnej wiedzy? Jeśli to nazwiemy

że autentycznej wiedzy nie można osiągnąć na podstawie doświadczenia zmysłowego. Idee bowiem, jako „wszczepione przez Boga w każdy umysł (…) *nie są* uogólnieniem doświadczenia zmysłowego”[31]. Nie są też jednak stale obecne w umyśle. Można właściwie stwierdzić, że według Kartezjusza, wrodzona jest tylko umysłowa *zdolność* wytwarzania tych idei.

Redukcjonizm Kartezjusza opierał się na uznaniu wszystkich ciał (niebieskich i ziemskich) za podpadające pod prawa mechaniki. Dotyczyło to również fizjologii, choć nie było jednoznaczne z uznaniem człowieka za maszynę. Maszynami były dla Kartezjusza zwierzęta, gdyż nie mają one duszy. Ludzkie ciało to z tego punktu widzenia również mechanizm, ale jego część, a mianowicie fragment mózgu zwany szyszynką, został przez Kartezjusza uznany za „siedlisko duszy”, „zmysł wspólny”[32] (*sensus communis*), w którym łączą się dane dostarczane przez wszystkie wyspecjalizowane zmysły, co umożliwia wzajemnie oddziaływanie na siebie umysłu i ciała[33]. Człowiek bez „uwięzionej” w szyszynce duszy, bez tego niematerialnego czynnika w materialnym ciele, nie byłby według Kartezjusza człowiekiem, lecz tylko „zwierzęcą maszyną”.

Wizja „duszy uwięzionej w maszynie”[34] do dziś jest zakorzeniona w myśleniu potocznym o człowieku, a kartezjańskie „wynalezienie umysłu”[35] miało wielki wpływ na europejską, a później amerykańską myśl naukową. Interakcyjny dualizm Kartezjusza okazał się jednak źródłem problemów, z którymi borykał się nie tylko jego twórca, ale również wielu późniejszych filozofów[36]. Descartes miał świadomość jedności umysłu i ciała w ich wzajemnym na siebie oddziaływaniu, co wobec uznania ich za jakościowo odrębne substancje, rodzi wiele kwestii dotyczących związków zachodzących między tymi substancjami[37].

---

przypominaniem sobie, słusznie to chyba nazwiemy” – Platon, *Fedon*, 75 E, w: Platon, *Dialogi*, t. I, tłum. W. Witwicki, Wyd. Antyk, Kęty 1999, s. 656.

[31] H.M. Bracken, *op. cit.*, s. 357.

[32] Pojęcie wprowadzone przez Arystotelesa w *O duszy*, III, 425 a, w: Arystoteles, *op. cit*., 1992, s. 111.

[33] Może to sugerować, że według Kartezjusza dusza (umysł) jest jakby ukrytym w mózgu *homunkulusem*.

[34] Jest to wizja bardzo podobna do tradycyjnego odczytu platońskiej wizji człowieka jako duszy, dla której ciało jest więzieniem. Paradoksalnie, element mechanicyzmu przywodzi z kolei na myśl wizję świata Demokryta (zaakceptowaną później przez Epikura), z którą ani Platon, ani Arystoteles zgodzić się nie mogli. Kartezjańska koncepcja budowy materii, jako substancji ciągłej, różniła się od starożytnej koncepcji materii o budowie ziarnistej.

[35] Określenie to, którego użył w opisie Kartezjusza m.in. Richard Rorty, zostało przytoczone przez Jerzego Bobryka w: *op. cit.*, s. 15.

[36] Dualizm kartezjański ma charakter nie tylko substancjalny, lecz również dotyczy stosunku zachodzącego między ciałem a umysłem. Kartezjusz głosił tezę wzajemnego oddziaływania na siebie tych dwóch substancji. Taki rodzaj dualizmu nazywa się dualizmem typu interakcyjnego. Innym dualizmem jest dualizm typu paralelnego. Aby ukazać różnicę między tymi dualizmami, posłużę się przykładem przedstawionym przez Davida Armstronga: „Dualizm typu interakcyjnego pojmuje stosunek między ciałem i umysłem analogicznie do stosunku między pokojem i termostatem. Ciało działa na umysł, a umysł oddziałuje na ciało. (…) Dualizm typu paralelnego pojmuje stosunek między ciałem i umysłem na wzór stosunku między pokojem i termometrem. Ciało działa na umysł, ale umysł w ogóle nie jest zdolny do oddziaływania na ciało. Istnieje jeszcze skrajniejsza forma paralelizmu, według której nie tylko umysł jest niezdolny do oddziaływania na ciało, ale i ciało jest niezdolne do oddziaływania na umysł. Po prostu istnieją one obok siebie jak (…) dwa doskonale zsynchronizowane zegary” – D. M. Armstrong, *Materialistyczna teoria umysłu*, tłum. H. Krahelska, PWN, Warszawa 1982, s. 14.

[37] Człowiek stanowi według Kartezjusza harmonijną jedność dwóch niezależnych od siebie substancji: myślącej i rozciągłej. Pytania pojawiły się w odniesieniu do tej harmonii.

1.2.2. Próby rozwiązania trudności dualizmu interakcyjnego

Próby rozwiązania kartezjańskiego problemu wzajemnej relacji między porządkiem cielesnym a duchowym prowadziły do formułowania nowych wizji świata i człowieka. Można tu wyróżnić dwa generalne kierunki – ten, który akceptuje i ten, który neguje dualizm psychofizyczny. W ich obrębie znajdują się cztery najważniejsze propozycje: okazjonalizm Nicolasa Malebranche'a (1638-1715), panteizm Barucha de Spinozy (1632-1677), monizm materialistyczny Tomasza Hobbesa (1588-1679) i monadologia Gottfrieda Wilhelma Leibniza (1646-1716).

1.2.2.1. Paralelizm psychofizyczny i monizm neutralny

Teoria Malebranche'a nie jest bezpośrednio związana z genealogią Sztucznej Inteligencji, jednak warta jest przytoczenia ze względu na fakt, że stanowi doskonały przykład paralelnego dualizmu psychofizycznego, który w odróżnieniu od dualizmu interakcyjnego, zaprzecza wzajemnemu oddziaływaniu duszy i ciała. Malebranche różnił się od Kartezjusza nie tylko w tej sprawie. Akceptował wprawdzie substancjalny dualizm psychofizyczny, ale odszedł od racjonalizmu i oparł się na podstawach teologicznych. Twierdził mianowicie, że możliwości przyczynowania są pozbawione nie tylko byty duchowe, ale także fizyczne. Nie pozostaje zatem nic innego, jak tylko przypisać jedyną moc sprawczą woli boskiej, uznając to, co nazywamy przyczyną, za *okazję*, którą Bóg może wykorzystać do dokonywania zmian w świecie zgodnie ze swoją wolą. Przyczyny rzeczywiste przerodziły się u Malebranche'a w przyczyny okazjonalne, a dostrzegana w świecie harmonia (przejawiająca się między innymi we współdziałaniu duszy i ciała) okazała się rezultatem interwencji boskiej, będącej wiecznym i ciągłym procesem.

Koncepcja Malebranche'a nigdy nie cieszyła się wielką popularnością. Warto jednak zwrócić uwagę, że z paralelizmem mamy do czynienia również w teorii Leibniza, którego przemyślenia w dużym stopniu wpłynęły na pojęcie umysłu w ideologii Sztucznej Inteligencji[38].

Zupełnie inną od zaproponowanej przez Malebranche'a wizję świata przedstawił Spinoza. Jest ona określana jako „panteistyczna", choć pojęcie „Boga Spinozy" znacznie odbiega od powszechnie przyjętych teistycznych definicji Boga osobowego.

Wyjaśnienie problemu psychofizycznego jest u Spinozy oparte na koncepcji Boga jako Natury, jako doskonałej i jedynej substancji[39]. Jako że nie jest to ani substancja materialna, ani duchowa, wizję Spinozy określa się mianem monizmu *neutralnego*. Ta radykalna orientacja monistyczna, choć swym racjonalizmem bliska Kartezjuszowi, istotnie różni ontologię spinozjańską od kartezjańskiej.

---

[38] Zob. podrozdział 1.2.2.3.

[39] Oto podstawowe definicje, sformułowane przez Spinozę w jego *Etyce*: „Przez substancję rozumiem to, co istnieje samo w sobie i pojmowane jest samo przez siebie, czyli to, czego pojęcie nie wymaga pojęcia innej rzeczy, za pomocą którego musiałoby być utworzone", „Przez atrybut rozumiem to, co rozum poznaje z substancji stanowiącej jej istotę", „Przez Boga rozumiem byt nieskończony bezwzględnie, to znaczy substancję składającą się z nieskończenie wielu atrybutów, z których każdy wyraża istotę wieczną i nieskończoną". Po sformułowaniu twierdzenia, że „w przyrodzie nie może być dwóch lub więcej substancji o tej samej naturze, czyli o tym samym atrybucie", Spinoza doszedł do wniosku, że „żadna substancja prócz Boga nie może ani istnieć, ani być pojęta" (cytaty zaczerpnięte z: T.L.S. Sprigge, *Spinoza* w: *Encyklopedia filozofii* pod red. T. Hondericha, tom II, Zysk i S-ka, Poznań 1998, s. 867).

Spinoza nie mówił o dwóch obcych sobie bytach, lecz o jedynym w pełni istniejącym bycie jako całości, której składniki nie mogą istnieć samodzielnie, lecz tylko „w pewnym stopniu”[40]. Jednym z takich skończonych i niesamodzielnych w swym istnieniu bytów jest człowiek, pojęty jako „samowiedza Natury”, „nośnik” świadomości.

W koncepcji Spinozy, Bóg jest „ideą siebie samego jako systemu fizycznego, i każda rzecz skończona (…) jest zarówno rzeczą fizyczną, jak i ideą, czyli tą składową bożego umysłu, który jest świadomością owej rzeczy”[41]. Wobec stwierdzenia możliwości ujęcia każdej rzeczy fizycznej jako idei tej rzeczy, Spinoza uznał ludzki umysł za ideę ludzkiego ciała jako funkcjonującej całości. Człowiek Spinozy, jako „istniejąca w Bogu idea ciała pobudzonego przez otoczenie”[42], działa na podstawie pojawiających się w umyśle idei otoczenia i dzięki nim samozachowawczo dostosowuje się to środowiska. W koncepcji spinozjańskiej każde zdarzenie ma swoją przyczynę w prawach przyrody, jako stałej naturze Boga, oraz w warunkach to zdarzenie poprzedzających. Jak zapewniał Spinoza, wynikającej z tego determinizmu niewoli można uniknąć tylko poprzez rozumowy wgląd intuicyjny[43].

Według Spinozy, Bóg jako jedyna substancja, posiada nieskończona ilość atrybutów, z których ludziom znane są tylko dwa: rozciągłość i myślenie. Filozof ten uznał, że kartezjańskie założenie istnienia dwóch substancji było nieuzasadnione, wobec czego dylemat wzajemnej relacji porządku duchowego i cielesnego po prostu nie istnieje. Jeśli bowiem substancja jest jedna, to rozciągłość i myślenie mogą być tylko jej atrybutami, których współzależność nie wymaga wytłumaczenia.

Do spinozjańskiej wizji świata nawiązał na początku XX wieku psycholog i filozof amerykański William James, który w rozprawie *Does Consciousness Exist?* przedstawił przyrodę jako składającą się z tworzywa, które może posiadać zarówno fizyczne, jak i psychiczne aspekty (atrybuty). Spinoza bywa obecnie uznawany za zwiastuna współczesnego podejścia do problemu relacji umysłu i ciała, opartego na dualizmie własności[44].

1.2.2.2. Monizm materialistyczny Tomasza Hobbesa

Tomasz Hobbes (1588-1679) odrzucił założenia kartezjańskiego dualizmu substancji, uznając substancję materialną za jedyną istniejącą. W swym monizmie Hobbes był tak samo

[40] Według takiej koncepcji pytanie, czy coś *istnieje* jest „nie na miejscu”, gdyż wszystko istnieje w różnym stopniu, partycypując w jedynym pełnym istnieniu Boga.
[41] T.L.S. Sprigge, *op.cit*, s. 868.
[42] *Ibidem*, s. 869.
[43] Według Spinozy, fizyczne i umysłowe zachowanie człowieka może wynikać z jego własnej istoty (jest to zachowanie *czynne*, w przypadku zachowania umysłowego zwane ideami *adekwatnymi*) albo może być powodowane czynnikami zewnętrznymi (jest to zachowanie *bierne*, w przypadku zachowania umysłowego zwane ideami *nieadekwatnymi*). Tylko idee adekwatne uważał Spinoza za składnik wiedzy rzetelnej, wiedzy najwyższego stopnia, najbardziej adekwatnej (Spinoza wyróżnił trzy stopnie wiedzy: wiedzę opartą na niejasnym doświadczeniu zmysłowym, wiedzę opartą na rozumowaniu i wiedzę uzyskaną dzięki intuicyjnemu wglądowi rozumowemu) – na podstawie: *Ibidem*.
[44] Można postawić zarzut Spinozie, że jego założenie istnienia jednej substancji jest tak samo nieuzasadnione, jak kartezjańskie założenie istnienia dwóch substancji. Niezaprzeczalną zaletą stanowiska Spinozy jest jednak to, że nie jest ono obarczone trudnościami dualizmu substancjalnego. Ale czy na pewno współzależność poszczególnych atrybutów substancji nie wymaga wyjaśnienia? Gdyby tak uważano, nigdy nie rozpoczęłaby się filozoficzna debata nad relacją własności fizycznych i mentalnych, która jest centralnym punktem rozważań prowadzonych w ramach współczesnej filozofii umysłu.

radykalny, jak Spinoza. Opisując stanowisko Hobbesa z punktu widzenia spinozjańskiej koncepcji świata, można stwierdzić, że monizm Hobbesa nie dotyczył substancji o nieskończonej liczbie atrybutów, lecz substancji posiadającej jedynie atrybuty materii, czyli właściwości geometryczne i mechaniczne. Taki pogląd pociągał za sobą uniwersalizm naukowy, który u Hobbesa przybrał postać mechanicyzmu[45].

Człowiek był dla Hobbesa mechaniczną strukturą, nieco bardziej skomplikowaną od struktury zwierzęcej. Zjawiska, których podłoże Descartes widział w substancji duchowej, Hobbes uznał za czysto fizyczne i podlegające prawom mechaniki. Podejście takie wynikało z determinizmu leżącego u podstaw jego filozofii[46]. Ruch każdego ciała to według Hobbesa konieczny efekt ruchu innych ciał, a założenie, że nic oprócz materii istnieć nie może, jest równoznaczne z negacją istnienia kierującej ludzkim ciałem substancji duchowej.

We wstępie do *Lewiatana* Hobbes porównywał wspólnotę ludzi do pojedynczego człowieka. Zawarł tam zdanie, które można uznać za zapowiedź wspartych funkcjonalizmem dwudziestowiecznych dążeń do stworzenia sztucznej inteligencji: „dlaczego nie moglibyśmy stwierdzić, że wszystkie automaty (silniki poruszające się dzięki sprężynom i kołom…) cechuje sztuczne życie? Bo czym jest serce, jeśli nie sprężyną; a nerwy, jeśli nie dużą ilością strun; a stawy, jeśli nie dużą ilością kół…”[47].

Hobbes twierdził, że myślenie ma swe źródło w doznaniach zmysłowych, które wywoływane są przez wywierany na oko nacisk cząsteczek, odbijających się od zewnętrznych ciał[48]. Rozumowanie było dla Hobbesa *obliczaniem* opartym na porównywalnych do reguł arytmetycznych prawach mechaniki. Filozof ten pisał, że „kiedy człowiek rozumuje, nie czyni nic innego, jak tylko pojmuje całość przez dodawanie części, gdyż rozumowanie (…) jest jedynie obliczaniem (…)”[49].

Jak zauważa Keith Devlin, przekonanie o możliwości sformalizowania wiedzy wynikło z przekładu greckiego słowa *logos*, oznaczającego nie tylko samo logiczne myślenie, ale i pojmowanie całych sytuacji. Słowo *logos* zostało jednak przełożone na łacińskie słowo *ratio*, które znaczy to samo, co słowo *obliczanie*. Stąd prawdopodobnie wynikło pojęciowe skojarzenie zwrotu „człowiek rozumny” ze zwrotem „człowiek obliczający”. Miało to niebagatelny wpływ na dalszy rozwój teorii umysłu.

Jak się później okaże, Hobbes (tak samo, jak późniejszy fizykalizm) nietrafnie zredukował wszystko do jednej dziedziny bytów. Rzeczywistość jest wprawdzie *jednością* i zasadnie można uznać za substancję coś fizycznego. Trzeba jednak pamiętać, iż nie oznacza to, że rzeczywistość jest *jednorodna*.

---

[45] Według Hobbesa, nauką uniwersalną, mogącą opisać całą rzeczywistość, nie powinna być tak bardzo szanowana przez Kartezjusza matematyka, lecz mechanika. Trudno się dziwić takiej postawie w okresie, kiedy to właśnie mechanika okazała się pierwszą nauką przyrodniczą, która osiągnęła ścisłą matematyczną formę.

[46] „Jakiekolwiek skutki już to powstaną, już to powstały, wszystkie one były konieczne ze względu na rzeczy, które je poprzedzały” – T. Hobbes, *O ciele*, II, 9, 5; E.F., I, tłum. C. Znamierowski, PWN, Warszawa 1956, s. 142 (za: *op. cit.*, R. H. Popkin (red.), s. 367).

[47] T. Hobbes, *Leviathan*, Wstęp, cytat za: Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/, (“why may we not say that all automata (engines that move themselves by springs and wheels…) have an artificial life? For what is the heart but a spring; and the nerves but so many strings; and the joints but so many wheels…”.

[48] Jest to wyraźny przejaw atomizmu, czyli pojmowania materii jako substancji ziarnistej, a nie ciągłej (jak u Kartezjusza). Widoczny jest tu także *mechanicyzm* (doznania są wywoływane mechanicznie).

[49] T. Hobbes, za: K. Devlin, *Żegnaj, Kartezjuszu*, tłum. B. Stanosz, Prószyński i S-ka, Warszawa 1999, s. 201.

1.2.2.3. Leibniz – monadologia i „maszyna myśląca”

Leibniz sprzeciwił się tak materializmowi Hobbesa, jak i monizmowi Spinozy, głosząc poglądy spirytualistyczne i pluralistyczne, natomiast w sprawie relacji duszy i ciała filozof ten był bliski paralelizmowi Malebranche’a.

Krytykę materializmu Leibniz oparł na racjonalnej analizie, która doprowadziła go do stwierdzenia, że rozciągłość i ekstensywność nie mogą być istotnymi właściwościami substancji, ponieważ to, co rozciągłe, musi być podzielne i jako takie nie może być czymś prostym. Wyprowadzona z tych przemyśleń teza, że może istnieć tylko substancja duchowa, dała początek koncepcji pluralistycznej, według której rzeczywistość to nieskończony zbiór substancji duchowych, zwanych przez Leibniza *monadami*.

Jeśli duchową monadę uzna się za substancję, to zgodnie z myślą Leibniza, ciało należy uważać za zjawisko tej substancji, za postać, „w jakiej jedna monada jawi się innej”[50]. Rozciągłość, uważana przez Kartezjusza za własność substancji materialnej, była dla Leibniza tylko względnym zjawiskiem postrzegającej monady. Postrzeżeniami Leibniz nazywał „stany o różnym stopniu jasności, wyrazistości i świadomości”[51]. Idąc za twierdzeniem, że monady reprezentują różny poziom rozwoju świadomości, Leibniz uważał duszę za monadę „wielce świadomą”, nie tylko postrzegającą, ale także zdolną do zapamiętania swych spostrzeżeń i korzystania z nich. Według Leibniza, dusza może się szczycić samowiedzą, którą posiada jeszcze tylko Bóg, jako monada najbardziej świadoma. Każdą monadę indywidualizuje zmienna treść świadomości, odzwierciedlająca ciągle zmieniającą się rzeczywistość[52]. Monady to uproszczone modele świata, mniej lub bardziej wiernie odzwierciedlające rzeczywistość *mikrokosmosy*[53]. Leibniz podkreślał niepowtarzalność monad i zakładał, że nie mogą one ulegać żadnym zewnętrznym oddziaływaniom (są *bezokienne*).

Wykluczenie percepcyjnego źródła poznania wymagało wyjaśnienia, jak może zachodzić zgodność między monadami a rzeczywistością. Wytłumaczenie tej zgodności Leibniz odnalazł w harmonii ustanowionej przez Boga przy tworzeniu świata[54].

Rzeczywistość była dla Leibniza *kosmosem*, harmonijną całością, czyli kolektywnym zbiorem uporządkowanych i powiązanych ze sobą bytów. Leibniz uważał, że harmonii tej nie tworzą wszystkie możliwe byty, lecz tylko takie, które mogą ze sobą współistnieć, czyli byty *współmożliwe*[55]. Bóg stworzył świat będący zbiorem takich bytów, świat „najlepszy z możliwych”, którego racjonalny i logiczny porządek wynika z „zaprogramowania monad”[56]. Tak rozumianą rzeczywistość Leibniz nazywał *Bożym Zegarem*, powstałym dzięki boskim obliczeniom.

---

[50] J. Bobryk, *op. cit.*, s. 38.
[51] W. Tatarkiewicz, *Historia filozofii*, t. II, PWN, Warszawa 1999, s. 78.
[52] Monady nie są więc odpowiednikiem *niezmiennych* idei platońskich.
[53] Jak twierdzi Jerzy Bobryk, „mikrokosmos-monada jest czymś izomorficznym, a nie podobnym do reszty rzeczywistości (…), w jakiś sposób ją reprezentuje logicznie. Każda monada jest inna, jest inną postacią reprezentacji, podobnie jak różne są matematyczne modele tej samej rzeczywistości.” – J. Bobryk, *op. cit.*, s. 19.
[54] Może się tu nasuwać wizja Boga jako wielkiego programisty, który wczytał w monady program reprezentujący rzeczywistość. Opis monady potraktowanej jako program byłby „tabelą działania” mechanizmu będącego jej przejawem.
[55] O tym, że nie wszystko może ze sobą współistnieć, świadczyć może choćby dobór naturalny, czy brak możliwości istnienia świata zbudowanego jednocześnie z cząstek i antycząstek.
[56] J. Bobryk, *op. cit.*, s. 38.

Według Leibniza, prawom mechaniki mogą podlegać tylko zjawiska, ale nie monady, które działają celowo, kierując się pożądaniem. Monadologia Leibniza zawiera wyzywające dla materialistycznej koncepcji umysłu stwierdzenie, w którym opis procesów mechanicznych okazuje się opisem co najmniej niewystarczającym dla wyjaśnienia tego, co dzieje się w umyśle. Leibniz pisze, że „postrzeżenie i to, co od niego zależy, nie da się wytłumaczyć racjami mechanicznymi" i gdybyśmy przyjęli, „że istnieje maszyna, której budowa pozwala, aby myślała, czuła, miewała postrzeżenia, będzie można pomyśleć ją, z zachowaniem tych samych proporcji, tak powiększoną, by można do niej wejść jak do młyna", a to założywszy, odnaleźlibyśmy w niej „tylko części, które popychają się wzajemnie, nigdy jednak nic, co tłumaczyłoby postrzeżenia"[57]. Dziś niektórzy zwolennicy metafory komputerowej zinterpretowaliby zapewne zacytowane wyżej zdanie następująco: w celu odkrycia tajemnic *software*'u (umysłu) nie należy ograniczać swych dążeń do obserwacji *hardware*'u (ciała), który jest tylko marną reprezentacją, niewyraźnym odbiciem, przejawem tego, co chcemy poznać. Ważny jest program, a nie komputer. Liczy się forma, a nie materia[58].

Leibniz był nie tylko filozofem, ale także zdolnym matematykiem[59]. Myślał o skonstruowaniu maszyny, która mogłaby wyręczyć człowieka w dokonywaniu obliczeń. Sporządził nawet plan budowy mechanicznego kalkulatora, ale nie udało mu się tego zamysłu zrealizować[60]. Nie zniechęciło to go do zajęcia się planem skonstruowania maszyny, która w drodze kalkulacji pomagałaby rozstrzygać spory filozoficzne[61] i mogłaby też służyć porozumiewaniu się ludzi mówiących różnymi językami[62]. Według Leibniza, podstawą funkcjonowania takiej maszyny musiałby być system językowy w pełni podlegający zasadom arytmetyki. Leibniz przedstawił opis takiego języka wewnętrznego maszyny, nazywając go *characteristica universalis*, zaś samą maszynę – *calculus universalis*. Pojęciom tego języka (będącym kombinacjami liter, symboli matematycznych i cyfr) towarzyszyć powinien odpowiedni słownik pojęciowy, obejmujący pojęcia podstawowe i wynikające z nich pojęcia pochodne. Leibniz twierdził, że język wewnętrzny tak zbudowanej maszyny liczącej musiałby stanowić systematyzację wiedzy, która byłaby podstawą wykonywanych operacji obliczeniowych. Uważał on, że cały proces poznawczy można ująć w system reguł, zastępując symboliką wszystkie działania umysłowe. Jeśli więc przedstawiona przez Leibniza maszyna posługiwałaby się opisanym wyżej językiem, to byłaby według niego maszyną myślącą[63].

---

[57] G.W. Leibniz, *Monadologia*, za: R. Kirk, *Mechanicyzm*, w: *op. cit.*, T. Hondericha (red.), t. II, s. 558.

[58] Można tutaj odwołać się do przykładu wydrukowanego w gazecie zdjęcia, które nie jest zdjęciem ze względu na papier i zaschnięte na nim kropki z farby drukarskiej. Zdjęcie jest zdjęciem dlatego, że kropki te, tworząc odpowiednią strukturę, są dla kogoś obrazem. Istotą zdjęcia jest jego forma. Ale czy tak, jak zdjęcie może być wywołane lub wydrukowane na różnych nośnikach, tak i umysł można zrealizować w dowolny sposób? Funkcjonalista dałby na to pytanie odpowiedź twierdzącą.

[59] Leibniz (niezależnie od Isaaca Newtona) odkrył rachunek różniczkowy. Dlatego rachunek ten często nazywa się *rachunkiem Leibniza-Newtona*.

[60] Zaproponowane przez Leibniza rozwiązania techniczne zostały wykorzystane w maszynach liczących używanych jeszcze w XX wieku.

[61] Filozofowie, rozstrzygający prawdziwość sądów (również semantyczną) za pomocą takiej maszyny, mieliby wołać przy tym *Calculemus!* (*Porachujmy!*) – stąd wzięła się nazwa maszyny kalkulującej Leibniza.

[62] Już Kartezjusz uważał, że główną przeszkodą „*w porozumiewaniu się ludzi mówiących różnymi językami (…)* [*jest*] *brak gramatyki uniwersalnej, skonstruowanej na zasadach logicznych, gramatyki pozbawionej wszelkich wyjątków, form nieregularnych i defektywnych.*" – M. Jurkowski, *Od wieży Babel do języka kosmitów. O językach sztucznych, uniwersalnych i międzynarodowych*, KAW, Białystok 1986, s. 22, za: M. J. Kasperski, *op.cit.*, s. 35-36.

[63] Oto cytat, który można by uznać za motto wielu prac ważnych dla Sztucznej Inteligencji (choćby rozpraw Jerry'ego A. Fodora, zawierających opisaną w dalszej części pracy koncepcję języka *mentaleskiego*): „Postęp sztuki inwencji rozumowej zależny jest w znacznej mierze od sztuki znakowania. (…) Gdyby utworzono jakiś ścisły język (nazwany przez niektórych Adamowym), albo przynajmniej pewnego rodzaju pismo prawdziwie

Calculus universalis można nazwać „maszyną myślącą w sensie Leibniza”[64]. Koncepcja tworzenia „myślących maszyn” jest współcześnie domeną mocnej odmiany Sztucznej Inteligencji. Nie jest chyba przesadą uznać Leibniza za najważniejszego nowożytnego prekursora tego nurtu. Bardzo ważne dla koncepcji Sztucznej Inteligencji jest widoczne zarówno u Leibniza, jak i u Hobbesa, kalkulacyjne ujęcie myślenia.

Idea zaproponowanej przez Leibniza *gramatyki uniwersalnej* pojawi się w tej pracy niejednokrotnie, zawsze w powiązaniu z obliczeniową koncepcją umysłu, tkwiącą u podstaw ideologii Sztucznej Inteligencji.

1.2.2.4. Mechanicyzm Juliana O. de La Mettrie

Rozprawa Juliana Offreya de La Mettrie, niedwuznacznie zatytułowana *Człowiek-maszyna*, miała stanowić kontynuację poglądów Kartezjusza, jednak bliższa jest twierdzeniom sformułowanym przez Tomasza Hobbesa; zawiera ona także krytyczne uwagi wobec *monadologii* Leibniza.

La Mettrie uważał się za zwolennika kartezjanizmu w twierdzeniu, że za maszynę należy uznać nie tylko zwierzę, ale także ludzką duszę, czyli samego człowieka. Podobnie, jak myśliciele starożytni, filozof ten pojmował duszę wielofunkcyjnie, jako substancję myślącą i ożywiającą ciało. Uważał przy tym, że źródłem ruchu ciała może być tylko inne ciało, co prowadzi do twierdzenia, że *dusza* jest substancją materialną. A skoro jedną z funkcji duszy jest myślenie, to musi być ono procesem mechanicznym. Mechanicyzm ten odbił swe piętno na współczesnej, fizykalistycznej teorii umysłu.

Już Kartezjusz dostrzegł zależność duszy od ciała, lecz La Mettrie posługiwał się obserwowalnymi dowodami tej zależności dla uargumentowania swych materialistycznych tez (warto zauważyć, że taka jest również droga argumentacji we współczesnym fizykalizmie i funkcjonalizmie).

La Mettrie, odwołując się do fizyki, chemii, fizjologii i anatomii, formułował twierdzenia, które wyrażają pogląd, że funkcje umysłowe są zależne od stanu centralnego układu nerwowego. Czyni to tego francuskiego lekarza nowożytnym prekursorem ważnego etapu we współczesnej filozofii umysłu, zwieńczonego tak zwaną *teorią identyczności stanów centralnych*.

---

filozoficzne, z pomocą którego pojęcia zostałyby sprowadzone do jakiegoś alfabetu myśli ludzkich, to by wszystko, do czego można dojść rozumem na podstawie danych, dało się uzyskać poprzez pewien swoisty rachunek w ten właśnie sposób, w jaki są rozwiązywane problemy arytmetyki lub geometrii” – G.W. Leibniz, *Die philosophischen Schriften von G.W. Leibniz*, tom VII, s. 198-199, przeł. M. Gordon, za: H. Święczkowska, *Algorytmiczność poznania według Leibniza*, w: *Jedność nauki – jedność świata?*, pod red. M. Hellena i J. Mączki, Biblos, Tarnów 2003, s. 179.

[64] Tak, jak to uczynił M. J. Kasperski w: *op. cit.*, s. 37.

1.3. Umysł w działaniu

1.3.1. Empiryzm Johna Locke'a

Wraz z filozofią Locke'a (1632-1704) argumentacja ontologiczna zaczęła ustępować miejsca argumentacji epistemologicznej. Locke zajął się problemem pochodzenia pojęć o bycie, a nie samym bytem, kierując uwagę na podmiot, a nie na przedmiot poznania. Nastąpiła tym samym antropologizacja filozofii.

Umysł pojmował Locke jako „niezapisaną tablicę", która nabiera treści tylko poprzez działanie, na drodze doświadczenia. Zgodnie z głoszonymi przez Locke'a tezami empiryzmu genetycznego, za jedyne wrodzone elementy umysłu należy uznać jego władze. Filozof ten nie był jednak sensualistą, gdyż wprowadził podział doświadczenia na zmysłowe „postrzeżenia" rzeczy zewnętrznych i umysłową „refleksję" (doświadczenie faktów wewnętrznych, czyli działania umysłu). „Zmysłom zewnętrznym" towarzyszył więc u Locke'a „zmysł wewnętrzny".

Locke łączył empiryzm w sprawie idei z racjonalizmem w stosunku do wiedzy, której same zmysły nie są w stanie dostarczyć. Według tego filozofa, wszystkie idee mają swe źródło w doświadczeniu i są tylko „materiałem dla rozumu i wiedzy". Wiedzę zaś pozwala zdobyć tylko rozum, wiążący ze sobą nabyte w drodze doświadczenia idee.

W przedmiotach spostrzeżeń Locke rozróżnił pierwotne jakości zmysłowe (*kształt* i *wielkość*) i wtórne jakości zmysłowe (takie, jak *zapach*, *smak*, *barwa*). Uznał przy tym, że źródłem powstania tych pierwszych może być kilka zmysłów, zaś w przypadku jakości wtórnych, jeden wyspecjalizowany zmysł. Według Locke'a, tylko jakości pierwotne mogą być podstawą wiedzy pewnej (tylko one mają charakter obiektywny). Jakości wtórne były dla Locke'a treściami świadomości, które odnosząc się do zmiennych cech przedmiotowych, mogą odzwierciedlać tylko stan podmiotu poznającego.

Locke wprowadził nowe pojęcie *idei*, nazywając nimi wszystko, co może się „znaleźć" w umyśle. W drodze dalszej analizy, Locke rozróżnił idee *proste*, pochodzące wyłącznie z doświadczenia, oraz idee *złożone* (…z idei prostych), które podzielił na idee *substancji*, *stosunków* i *objawów*[65].

Według Locke'a, doświadczenie nie może być źródłem wiedzy o substancjach, gdyż dotyczy jedynie własności (atrybutów). Locke uważał pojecie substancji za wytwór umysłu, który ma zaspokoić potrzeby niezmiennej „podpory" dla zmiennych cech (dla idei, które umysł musi „składać" w procesie myślenia). Filozof ten twierdził, że w miarę wzrostu złożoności idei, współczynnik dowolności odgrywa coraz większą rolę, pomniejszając pewność poznania. Skoro więc doświadczenie nie daje nam żadnej wiedzy na temat substancji, to jest ona według Locke'a niepoznawalna[66].

Niewielki stopień pewności Locke przypisywał nie tylko twierdzeniom metafizycznym, ale także tezom z dziedziny fizyki. Poznanie pewne w sferze idei złożonych było według Locke'a możliwe tylko w zakresie matematyki i etyki, co zbliżało „ojca nowożytnego empiryzmu" do

[65] W dalszej części tej pracy stanie się widoczne, że ślady koncepcji Locke'a można dostrzec między innymi w *psychofunkcjonalizmie* Jerry'ego Fodora.

[66] Locke nie wypowiadał się w sprawach ontologicznych, dotyczących istnienia i natury substancji. Stwierdzenie braku możliwości poznania substancji nie jest równoznaczne z zaprzeczeniem jej istnienia.

racjonalizmu w duchu Kartezjusza. Trudno się temu dziwić, jeśli weźmie się pod uwagę, że poznanie było dla Locke'a intuicyjnym „postrzeżeniem związku i zgodności albo niezgodności i przeciwieństwa między naszymi ideami"[67]. Można wobec tego za pomocą „refleksji" ustalić zgodność idei (możemy po prostu wiedzieć, że 2+2=4), ale o tym, czy idee te rzeczywiście *coś* reprezentują, można dowiedzieć się tylko w drodze obserwacji przedmiotów jednostkowych.

Teoria „zmysłu wewnętrznego" będzie później stanowiła podstawę dla psychologii *introspektywnej*, z którą toczył walkę obiektywistyczny behawioryzm, „spadkodawca" funkcjonalistycznej teorii umysłu. Nie były to jedyne wpływy empiryzmu Locke'a na psychologię, która dzięki jego następcom, traktującym psychikę mechanistycznie, wzbogaciła się o *asocjacjonizm*.

Z kolei funkcjonalizm[68] zawdzięcza Locke'owi koncepcję umysłu funkcjonującego w oparciu o *przedstawienia* przedmiotów i ich cech. Locke twierdził, że żadna z rzeczy rozważanych w umyśle *nie jest* dana bezpośrednio, z czego wynika, że musi istnieć jakaś inna rzecz, która *jest* dana bezpośrednio, jako *reprezentacja* rozważanej rzeczy. Takimi reprezentacjami, czyli znakami rzeczy badanych w umyśle są według Locke'a idee. *Reprezentacjonizm* ten wynika z tej części *Rozważań dotyczących rozumu ludzkiego*, której tematem jest język[69].

Język był dla Locke'a narzędziem służącym ludziom do wymiany myśli, które jako idee, wyrażane są w postaci przypisanych im znaków (na przykład dźwięków artykułowanych). Badanie tego lingwistycznego narzędzia było według Locke'a jednym ze sposobów osiągnięcia wiedzy na temat ludzkiego poznania. Locke twierdził, że „słowa w swoim bezpośrednim znaczeniu odpowiadają ideom tego, kto się nimi posługuje. W przeciwnym wypadku byłyby tylko pustymi dźwiękami. Aby jednak funkcjonować w komunikacji międzyludzkiej, muszą też wywoływać odpowiednie idee w słuchaczu"[70].

Dla potrzeb tej pracy warto zwrócić uwagę na to, że Locke dostrzegł trójczłonowość relacji poznawczej, to jest, że musi ona być zapośredniczona przez treść poznawczą (ideę, znak, reprezentację), która jest bezpośrednio dostępna podmiotowi (choć to nie jej dotyczy bezpośrednio akt poznania). Znaki są dane bezpośrednio w refleksji, jako reprezentujące coś, co nie jest dane bezpośrednio. Podmiot nie poznaje jednak samych znaków (idei), lecz to, co one znaczą, oznaczają, czy wskazują (reprezentują)[71]. Ta trójczłonowość jest podstawą reprezentacjonizmu, leżącego u podstaw psychofunkcjonalizmu.

1.3.2. Asocjacjonizm Davida Hume'a

Idąc śladami Locke'a, choć zmieniając nieco terminologię, David Hume podzielił przedstawienia (reprezentacje) na pierwotne *wrażenia* i pochodne od nich *idee*, które są

[67] J. Locke, *Rozważania dotyczące rozumu ludzkiego*, przeł. B.J. Gawędzki, PWN, Warszawa 1995, t. II, s. 194 za: G.A.J. Rogers, *John Locke*, w: *op. cit.*, R.H. Popkin (red.), s. 401.
[68] W szczególności funkcjonalizm Jerry'ego A. Fodora.
[69] Tę semiotyczną część swych *Rozważań…* Locke zatytułował „O słowach".
[70] Za pomocą słów nie można jednak poznać rzeczy samych w sobie, a to z powodu braku pewności, czy idee reprezentują te rzeczy takimi, jakimi naprawdę są. Jest to jeden ze zidentyfikowanych przez Locke'a problemów, których rozwiązaniem powinna się zająć teoria poznania. Cytat: G.A.J. Rogers, *John Locke*, w: *op. cit.*, R.H. Popkin (red.), s. 401.
[71] Locke nie wziął pod uwagę tego, że jeśli idee coś reprezentują i są przez to znakami, to muszą być takie, jak znaki – *przeźroczyste*.

wytwarzanymi przez umysł kopiami wrażeń. Wartość poznawczą idei Hume uzależnił od wierności kopiowania wrażeń. Wiedza może dotyczyć według Hume'a bądź faktów (poznawanych za pomocą wrażeń), bądź stosunków między ideami. Badanie stosunków między ideami odnosi się na przykład do twierdzeń matematycznych, których pewność jest niezależna od doświadczenia. Pewności takiej nie przypisał Hume twierdzeniom o faktach, co zmusiło go do szukania odpowiedzi na pytanie, czy poznanie może wykraczać poza stwierdzone fakty.

W toku swych rozważań Hume stwierdził, że wyjście poza stwierdzone fakty ma miejsce wtedy, gdy dopiero spodziewamy się zaistnienia jakichś faktów, tłumacząc to zauważonym wcześniej związkiem kauzalnym miedzy faktami. Uznając, że związki przyczynowo-skutkowe opierają się na instynktownym przyzwyczajeniu, czyli wrodzonej skłonności umysłu do oczekiwania powtórzeń, Hume nie tylko zakwestionował wartość poznawczą zasady przyczynowości, ale również zbudował fundamenty *asocjacjonizmu*[72]. Na podobnej zasadzie Hume kwestionował istnienie substancji, której przyjęcie było według niego nieuzasadnionym, instynktownym wykroczeniem poza fakty, mającym swe podłoże w naturalnych skłonnościach umysłu.

Trudno nie odnieść wrażenia, że przemyślenia Hume'a były w dużej części komentarzami do filozofii Locke'a, konsekwentnie utrzymywanymi z dala od twierdzeń ontologicznych. Hume różnił się nadto od Locke'a w kwestii poznania rzeczywistości, gdzie rozum zastąpił biologicznie uwarunkowanym instynktem.

1.4. Współczesny materializm redukcjonistyczny

1.4.1. Behawioryzm

1.4.1.1. Behawioryzm psychologiczny (metodologiczny)

W roku 1861 Paul Broca odkrył, że uszkodzenie pewnych obszarów ludzkiego mózgu ma wpływ na zdolność mówienia. Dziewięć lat później Gustav Fritsch postanowił zbadać dokładnie funkcje mózgu. Dokonując eksperymentów na psie, zaobserwował, że pobudzenie impulsem elektrycznym poszczególnych części mózgu skutkuje kurczeniem się określonych mięśni. Były to tylko niektóre z naukowych potwierdzeń tezy, że ciało i dusza[73] są ze sobą ściśle związane, że nie są nawet odmiennymi substancjami, co w pewnym stopniu tłumaczy ich wzajemny na siebie wpływ. Ponieważ funkcjonujący mózg jest częścią żywego organizmu, w psychologii zaczął się kształtować biologiczny punkt widzenia, którego współczesnym przejawem jest *behawioryzm*[74]. Behawioryzm można uznać za teorię materialistyczną, choć sama psychologia narodziła się z dala od ontologii. Powstał on jako alternatywa dla psychologii introspekcyjnej.

[72] Warto zwrócić uwagę, że Hume nie zaprzeczał przy tym istnieniu związków przyczynowych w realnym świecie. Uważał je tylko za niepoznawalne.
[73] Dusza rozumiana zarówno jako starożytna zasada ruchu (życia), jak i kartezjańska substancja myśląca.
[74] W 1986 roku psycholog Richard Thompson stwierdził, że „jedyną możliwością wyjaśnienia podstawowych problemów psychologii – jak choćby istoty doznawania, spostrzegania, pamięci i myślenia – jest zrozumienie biologicznego nośnika, to znaczy mózgu i jego procesów. (…) Jest to sfera psychologii." – cytat za: G. Mietzel, *Wprowadzenie do psychologii*, tłum. E. Pankiewicz, GWP, Gdańsk 1998, s. 25.

Pierwszy na świecie instytut psychologiczny został założony przez Wilhelma Wundta w Lipsku, w 1897 roku. Badania prowadzone w tym ośrodku dały początek psychologii doświadczalnej. Wundt kierował się w swych badaniach nadzieją, że dzięki introspekcji, czyli słownym sprawozdaniom o wrażeniach, uczuciach i wyobrażeniach, uda mu się dogłębnie poznać ludzką psychikę[75].

W roku 1913 amerykański psycholog John Watson (1878-1958) uznał metodę Wundta za nienaukową i przedstawił własny program badawczy, zwany behawioryzmem[76]. Nie był to jeszcze program wkraczający w dziedzinę badań biologicznych. Był to jednak początek tradycji, która dominowała w psychologii anglosaskiej aż do lat sześćdziesiątych XX wieku i ma swoich zwolenników do dziś[77].

Naukowość behawioryzmu opierała się na porzuceniu w badaniach punktu widzenia pierwszej osoby (introspekcji) i ograniczeniu się do przedmiotów badań dostępnych wielu niezależnym obserwatorom[78]. Na wzór nauk przyrodniczych, Watson postanowił skierować uwagę na badanie tego, co każdy może sprawdzić, a więc na zachowanie i jego uwarunkowania[79]. Bodźce i następujące po nich reakcje są przecież zauważalne i wymierne. Argumentem były tutaj eksperymenty rosyjskiego fizjologa Iwana Pawłowa, który wyjaśnił zachowanie psa, opisując tylko reakcje na poszczególne bodźce, zupełnie przy tym pomijając procesy zachodzące w organizmie zwierzęcia.

Według Watsona, w psychologii nie powinno się używać pojęć niejasnych, dotyczących doznań psychicznych, które występują w ogarniętym mrokiem umyśle, nazwanym przez uczniów Watsona „czarną skrzynką”[80]. Przyczyn zachowania Watson nie upatrywał we wrodzonych instynktach, lecz w wynikach wcześniejszego uczenia się. Traktując człowieka jako „niezapisaną tablicę”, psycholog ten starał się wykazać, że ludzkie zachowanie zależne jest wyłącznie od bodźców środowiskowych.

W określeniu metodologii behawiorystycznej zwykle cytuje się twierdzenia Burrhusa Skinnera (1904-1990), który pod wpływem poglądów Watsona uznał naukowa rzetelność psychologii, jako nauki o zachowaniach, wymaga odżegnania się od mówienia o subiektywnych zjawiskach psychicznych. Zjawiska psychiczne były dla Skinnera tylko przykładami behawioralnych reakcji na bodźce zewnętrzne. Za jedyne źródło zachowania uważał Skinner środowisko, a nie jakieś myśli, czy wyobrażenia. Nie bez powodu więc uważa się Skinnera za pioniera tak zwanego *behawioryzmu radykalnego*, według którego zachowaniem człowieka można kierować, gdyż jest ono zależne wyłącznie od bodźców

[75] Warto przypomnieć tu koncepcję zmysłu wewnętrznego Johna Locke’a, którą można uznać za filozoficzną podstawę badań opartych na introspekcji.

[76] Tezę o nienaukowości introspekcji Watson powtórzył potem w swej najbardziej wpływowej pracy stwierdzając, że „stany świadomości, podobnie jak tzw. zjawiska spirytystyczne, nie dają się obiektywnie zweryfikować i z tego powodu nigdy nie mogą stać się danymi naukowymi” – J. Watson, *Psychologia ze stanowiska behawiorysty*, 1919, s. 1, za: P.G. Zimbardo, *Psychologia i życie*, tłum. J. Łuczyński, PWN, Warszawa 1999, s. 313.

[77] Najbardziej skuteczną krytykę behawioryzmu przedstawił w 1959 Noam Chomsky.

[78] W dalszej części pracy ukażę, że porzucenie punktu widzenia pierwszej osoby zaowocowało zasadniczymi problemami, z jakimi musiał się potem borykać funkcjonalizm i wraz z nim teoretycy Sztucznej Inteligencji.

[79] „Chcemy ograniczyć się do spraw, które dają się zaobserwować (…). Ale co możemy zaobserwować? Zachowanie, to, co organizm czyni i mówi” – J. Watson, *Behaviorism*, 1925, za: G. Mietzel, *op. cit.*, s. 30.

[80] Do dziś mianem „czarnej skrzynki” nazywa się system o nieznanej i nieistotnej zasadzie działania. Termin ten znalazł swe zastosowanie szczególnie w funkcjonalistycznej teorii umysłu, o której piszę szerzej w rozdziale drugim.

zewnętrznych. Skinner sugerował, że ludzie i zwierzęta działają jak maszyny, które automatycznie reagują na określone bodźce[81].

We współczesnym behawioryzmie mamy do czynienia z redukowaniem tego, co psychiczne do tego, co fizyczne. Redukcji tej dokonuje się na mocy empirycznych praw naukowych, opartych na intersubiektywnie obserwowalnych zjawiskach, jakimi są zachowania. Za autorów takiego redukcjonistycznego podejścia uważa się najczęściej Herberta Feigla i Johna Smarta.

Smart stwierdził, że „«doznanie» i «proces w mózgu» są dwiema różnymi nazwami jednego i tego samego”[82] – tak, jak „błyskawica” odpowiada „wyładowaniu elektrycznemu”. Mamy tu do czynienia z wypowiedzią języka potocznego z jednej strony, a z drugiej strony, z wypowiedzią wywodzącą się z jakiejś teorii naukowej. Według Smarta obydwie wypowiedzi mają wprawdzie różny sens (*meaning*), ale posiadają to samo odniesienie przedmiotowe (*reference*), które można scharakteryzować w naukowych kategoriach fizykalnych.

Następstwem tez behawiorystycznych, skonfrontowanych z rozwojem neurofizjologii, była tak zwana „teoria identyczności rodzaju (*type-type*)”[83], którą będę nazywał „teorią identyczności”. Tę koncepcję „powszechnego fizykalizmu”[84], można uznać za słabszą wersję teorii identyczności egzemplarzy, gdyż stany psychiczne i stany neurofizjologiczne są w niej utożsamiane ze sobą nie tylko egzemplarycznie, ale także pod względem ich typu[85].

Kontynuatorami koncepcji psychofizycznej identyczności zostali między innymi David Armstrong i David Lewis[86] oraz Paul M. Churchland i Patricia Churchland. Jednym z poważnych sprzeciwów wobec tej teorii był sformułowany przez Hilarego Putnama argument z wielorakiej realizacji, zapowiadający nadejście nowej, funkcjonalistycznej koncepcji umysłu[87].

1.4.1.2. Behawioryzm logiczny

Reakcją na słabe punkty *teorii analogiczności*[88] oraz próbą wyjaśnienia relacji miedzy bodźcami a zachowaniem, był *behawioryzm logiczny*, z którym wiąże się Gilberta Ryle’a. Ryle określił dualizm kartezjański *pomyłką kategorialną*, polegającą na zaliczaniu umysłu i ciała do tej samej kategorii logicznej i przedstawianiu różnic pomiędzy nimi za pomocą jednej aparatury pojęciowej. Z sugestii Ryle’a wynika, że stwierdzenia o umyśle są tylko wypowiedziami na temat ciał, formułowanymi w języku potocznym. W rezultacie swych

[81] Podobnie myślał o organizmach La Mettrie.
[82] *Ibidem*, s. 97.
[83] Za: J. Bremer, *op. cit.*, s. 94.
[84] Teorię tę czasami nazywa się także *fizykalizmem typu* – za: *ibidem*, s. 94.
[85] Postuluje się tu tezę o identyczności nie tylko poszczególnych *egzemplarzy* stanów psychicznych, ale także tezę o identyczności stanów psychicznych danego *typu*.
[86] D. Armstrong i D. Lewis uznawani są za przedstawicieli *słabego* funkcjonalizmu, opartego na teorii identyczności egzemplarzy.
[87] H. Putnam, jako współtwórca *mocnego* funkcjonalizmu, otwarcie odrzucał teorię identyczności.
[88] Zgodnie z teorią analogiczności, ustalając korelacje pomiędzy własnymi stanami mentalnymi a zachowaniami, przez analogię stwierdzamy stany mentalne u innych na podstawie ich zachowań. Wymaga to założenia, że związki zachodzące pomiędzy zjawiskami umysłowymi, a zachowaniami, są u wszystkich osób takie same. Jak zauważa Józef Bremer, według krytyków teorii analogiczności, jednym z jej słabych punktów jest to, że nie ma możliwości „sprawdzenia, czy u wszystkich ludzi te same zachowania są zawsze związane z tymi samymi stanami mentalnymi” – J. Bremer, *op. cit.*, s. 57.

przemyśleń na temat tak pojętego dualizmu językowego, Ryle uznał, że umysł można sensownie opisywać tylko w odniesieniu do ludzkich zdolności i skłonności do zachowań. Z takiego punktu widzenia to, co dzieje się w umyśle, ma swoja podstawę w *dyspozycjach* do określonych zachowań, w tym użycia języka potocznego[89].

Behawioryzm logiczny, jako teoria semantyczna, postulował uznanie zdań niesprawdzalnych intersubiektywnie (empirycznie) za nienaukowe i wobec tego bezsensowne. Według tej odmiany behawioryzmu sensowne są tylko wypowiedzi o zjawiskach umysłowych analitycznie sprowadzalne do zdań o zachowaniach. Sens ma tylko takie pojęcie psychologiczne, które da się „przetłumaczyć" na język fizyki. Koncepcja ta stanowi analityczne rozwinięcie wymogu naukowej rzetelności psychologii.

Behawioryzm logiczny, na mocy swych definicji, wyeliminował ze swojego słownika terminy mentalistyczne, określające w języku potocznym tak zwane stany umysłowe. Według behawioryzmu logicznego, wypowiedzi o poszczególnych egzemplarzach takich stanów są niczym więcej, jak tylko złożonymi wypowiedziami o zachowaniach (w tym językowych) jako egzemplarzach zdarzeń.

1.5. Język i obliczanie

1.5.1. Psychologia kognitywna

Decydujący wpływ na przezwyciężenie behawiorystycznego wymogu rezygnacji z badania procesów zachodzących w organizmie miał rozwój techniki. Na początku lat pięćdziesiątych XX wieku wprowadzono do powszechnego użytku elektroniczne maszyny liczące, które po odpowiednim zaprogramowaniu, wskutek przetworzenia dostarczonych informacji, mogły wyręczyć człowieka w rozwiązywaniu niektórych problemów. Sądzono, że ludzie (tak, jak te maszyny) odbierają dane, przetwarzają je i zachowują w pamięci, aby w razie potrzeby mieć do nich dostęp. Wielu psychologów przestało w związku z tym zajmować się bodźcami jako faktami fizycznymi, poszukując w zamian odpowiedzi na pytania, *jak* człowiek te bodźce odbiera, interpretuje, zapamiętuje i łączy. Centralnym punktem zainteresowań stał się ludzki proces poznawczy, którego aktywność kłóciła się z nakreślonym przez behawiorystów obrazem człowieka. Tak narodził się nurt *psychologii kognitywnej*, który kwestionował ograniczone perspektywy behawioryzmu, nie przywiązującego żadnej wagi do kwestii związanych z myśleniem, czy świadomością.

Tak zwana „rewolucja kognitywna" trwa w psychologii od połowy lat sześćdziesiątych. Według psychologów kognitywnych, zachowanie jednostki ludzkiej jest zdeterminowane nie tylko przez poprzedzające je bodźce i konsekwencje wcześniejszych na te bodźce reakcji. We

[89] Ważny jest fakt, że behawioryzm logiczny opisuje umysł w kategoriach dyspozycji do zachowania, a nie samego zachowania. Pozwala to twierdzić, że inteligentne może być nawet coś, co nie ma szansy udowodnienia swej inteligencji poprzez odpowiednie behawioralne reakcje na bodźce. Jest to ujęcie nacechowane swoistym liberalizmem, ponieważ jest bliskie sugestii, że inteligentne może być wszystko. Jeśli uznać to za jedną z podstaw obrony ideologii Sztucznej Inteligencji, to jest to podstawa bardzo złudna. Tylko pozornie unika się tu twierdzeń o byciu inteligentnym, formułowanych tylko na podstawie obserwowanego zachowania. Jak bowiem inaczej, niż poprzez obserwację zachowania, można stwierdzić, czy w czymkolwiek (lub kimkolwiek) występują wyżej wspomniane dyspozycje? Sedno w tym, że inteligencja nie sprowadza się do samych bodźców i (uwarunkowanych dyspozycjami behawioralnymi) reakcji na te bodźce. Temat inteligencji powróci jeszcze w ostatnim rozdziale tej pracy.

współczesnej psychologii kognitywnej zakłada się istnienie tak zwanych „matryc umysłowych", czyli wrodzonego potencjału, który może być wykorzystany w różny sposób, owocując różnymi typami osobowości[90].

Zakładając bezpośredni związek umysłu z mózgiem, prowadzi się obecnie badania procesów zachodzących w mózgu na poziomie mikroskopowym (analizując na przykład działanie mózgu podczas wykonywania określonych zadań poznawczych)[91] i makroskopowym (śledząc na przykład wpływ zdarzeń z dzieciństwa na osobowość)[92]. Bierze się przy tym pod uwagę sposób interpretacji środowiska bodźcowego, uważając myśli za przyczyny zachowania[93].

Kognitywizm wygląda więc obiecująco, ale dziedziczy po fizykalizmie identyfikację umysłu z mózgiem, a po funkcjonalizmie tak zwaną *metaforę komputerową*, co według kognitywistów pociąga za sobą możliwość symulowania umysłowych procesów poznawczych za pomocą tworzonych komputerowo sztucznych sieci neuronowych. Na temat konsekwencji takiego podejścia piszę w rozdziale trzecim.

1.5.2. Noam Chomsky i gramatyka generatywna

Rok 1957 był dla Noama Chomsky'ego przełomowy ze względu na publikację dzieła *Syntactic Structures*, które przyniosło temu amerykańskiemu lingwiście wielki rozgłos. W książce tej Chomsky przedstawił gramatykę jako „pewnego rodzaju urządzenie do produkowania zdań języka"[94]. Nie chodziło mu przy tym o pojmowanie gramatyki jako elektronicznej lub mechanicznej maszyny naśladującej zachowanie człowieka mówiącego jakimś językiem. Należy tu raczej odwołać się do formalizacji gramatyki z wykorzystaniem takich słów, jak „urządzenie", czy „maszyna" w sensie pojęć matematycznych, mających znaczenie abstrakcyjne, nie związane z jakąkolwiek realizacja fizyczną.

„Produkowanie" zdań odbywa się według Chomsky'ego poprzez stosowanie ściśle ustalonych reguł, ale gramatyka odnosi się nie tylko do nich, lecz ma również własność warunkującą twórczy charakter języka. Chomsky twierdził, że większość zdań należących do każdego zbioru zarejestrowanych wypowiedzi to zdania nowe, występujące tylko jeden raz. Na pytania dotyczące zdolności użytkowników języka do rozumienia i produkowania zdań wcześniej niesłyszanych, Chomsky odpowiada, że gramatyka *generuje* zdania danego języka, określając je jako gramatyczne bez względu na to, czy były już użyte, czy nie. Z tego punktu widzenia, język można zdefiniować jako leksykon i zbiór wszystkich zdań generowanych przez gramatykę.

Chomsky przyjął, że wobec braku możliwości wskazania ograniczeń długości zdań języka angielskiego, ich zbiór należy uznać za nieskończony, nawet przy założeniu niezmienności i skończoności słownika tego języka oraz ograniczeniu ilości możliwych operacji związanych z

[90] Według psychologii kognitywnej człowiek rodzi się jako *tabula rasa*, jednak z „wyposażeniem" potrzebnym dla wielokierunkowego rozwoju, którego efekty są zależne od wielu czynników zewnętrznych i wewnętrznych.
[91] Sądzę, że zarówno Gustav Fritsch, jak i Wilhelm Wundt, z wielką satysfakcją i zapałem uczestniczyliby w obecnie prowadzonych badaniach psychologicznych, korzystając jednocześnie z analiz dokonanych przez Watsona i Skinnera.
[92] Poziom analiz makroskopowych wykorzystywał w swych badaniach na przykład Zygmunt Freud (1856 – 1939).
[93] Nie bez wpływu na psychologię kognitywną pozostaje filozofia Locke'a, który uważał, że wrodzone człowiekowi są tylko władze umysłu, a samo poznanie zależy od zmysłów.
[94] Cytat za: J. Lyons, *Chomsky*, Prószyński i S-ka, Warszawa 1998, s. 47.

generowaniem zdań. Ograniczenie to ukazuje, że można ustalić skończony zbiór reguł, za pomocą którego generowane są wszystkie zdania danego języka. Jeśli reguły te stosuje się do generowania możliwie nieskończonej liczby zdań w oparciu o skończony słownik, to w zbiorze tych reguł muszą znajdować się reguły rekursywne, dające się zastosować wiele razy przy generowaniu tego samego zdania. Tak, jak algebraiczna teoria rekursji[95] dała początek komputerystyce jako nauce o obliczeniach, tak teoria struktur syntaktycznych Chomsky'ego stanowiła początek nowoczesnej lingwistyki jako quasi-matematycznej nauki o języku.

Chomsky przedstawił w *Syntactic Structures* koncepcję gramatyki czysto syntaktycznej, opartej na tezie o autonomii składni i założeniu, że w konstruowaniu lub analizie takiej gramatyki można abstrahować od semantyki. Wobec zarzutu, że teza o autonomii składni jest równoznaczna z nawoływaniem do wyeliminowania semantyki z teorii lingwistycznej, Chomsky zaznaczył, iż wskazuje ona jedynie na to, że przy badaniach formy języka (syntaksy) nie trzeba odwoływać się do treści, ponieważ znaczenie semantyczne nie stanowi kryterium gramatyczności. Pojawił się jednak problem relacji między składnią a semantyką, będący konsekwencją faktu, iż teza o autonomii składni doprowadziła do uznania zanalizowanej i opracowanej gramatycznie syntaksy za użyteczną dla opisu semantycznego.

Chomsky, idąc za tradycyjnym dla lingwistyki strukturalnej rozróżnieniem *langue* i *parole*[96], odróżnił *kompetencję* językową od *wykonania*. Uważał jednak, że formowanie zdań należy do języka, a nie do mowy, że języka nie powinno się opisywać pragmatycznie (jak czynił to de Saussure), lecz jako generujący zdania system reguł. Chomsky utrzymywał, że podlegający matematyzacji opis kompetencji językowej jest możliwie najlepszym opisem dla celów teorii lingwistycznej, chociaż stanowi pewną idealizację.

W toku swych dalszych przemyśleń Chomsky stwierdził, że lingwistyka ma ogromne znaczenie w rozwiązaniu filozoficznego sporu między racjonalizmem a empiryzmem, którym nacechowane są treści zawarte w *Syntactic Structures*. Według Kartezjusza, percepcja i rozumienie świata wymaga wrodzonej, niezależnej od doświadczenia, zdolności do tworzenia idei. Empiryści brytyjscy (Locke, Berkeley i Hume) utrzymywali, że poznanie można sprowadzić do biernego rejestrowania wrażeń zmysłowych i kojarzenia ich na zasadzie asocjacji, co prowadziło do zatarcia różnicy między człowiekiem a zwierzęciem. Pod wpływem empiryzmu, fizykalizmu[97] i determinizmu[98], ukształtował się w psychologii pogląd, że wiedza i zachowanie człowieka są zdeterminowane wyłącznie przez środowisko.

Taką fizykalistyczno-deterministyczną koncepcją jest behawioryzm, redukujący umysł do dyspozycji do zachowania i zaprzeczający możliwości przyczynowego oddziaływania umysłu na ciało, czego uzasadnieniem miało być stwierdzenie, że to, co nazywamy umysłem, wiąże się z potocznymi, nienaukowymi wierzeniami i przekonaniami. Chomsky przyczynił się do upadku behawioryzmu, publikując w 1959 roku krytyczną recenzję pracy Burrhusa F. Skinnera, zatytułowanej *Verbal Behavior*.

[95] Chodzi tu o algebrę obliczeń zbudowaną w pierwszej połowie XX wieku przez Kurta Gödela, Alana Turinga, Alonzo Churcha i Stephena Kleene'a.
[96] Rozróżnienie *langue* i *parole* zostanie omówione w rozdziale trzecim.
[97] Fizykalizm pozwala na takie przeformułowanie twierdzeń o doznaniach, myślach i uczuciach, że stają się one twierdzeniami, które dotyczą zjawisk podlegających prawom fizycznym. Twierdzenia takie odnoszą się do stanów ciał i obserwowalnego zachowania.
[98] Zgodnie z doktryną determinizmu wszelkie zjawiska i zdarzenia fizyczne podlegają prawom przyczynowym, są zdeterminowane przez poprzedzające je inne zjawiska i zdarzenia fizyczne.

Podstawowym zarzutem sformułowanym przez Chomsky'ego wobec poglądów behawiorystycznych było wynikające z nich oparcie teorii uczenia się na zasadach asocjacji, podczas gdy według Chomsky'ego podstawą tej teorii jest zagadnienie kompetencji językowej. Wykazując wady opisu behawiorystycznego, Chomsky „zaproponował opis zjawisk językowych, który jest właściwie specyficzną wersją strukturalizmu"[99].

Uznając problemy *semantyki* za należące do wspólnego obszaru zainteresowań lingwistyki, psychologii i filozofii, Chomsky przyczynił się do narodzin *psycholingwistyki*, mającej na celu rozwiązanie tych problemów poprzez badanie cech i zasad funkcjonowania ludzkiego umysłu. Skupiono się przy tym na analizie semantycznych składników pamięci trwałej człowieka[100]. Analiza ta doprowadziła psycholingwistów do sformułowania teorii *pierwotnych* składników semantycznych, które jako elementarne uznano za wrodzone i zależne od ludzkich cech biologicznych[101]. Z czasem zarzucono tę koncepcję ze względu na stwierdzenie fałszywości domniemania, iż system poznawczy człowieka da się opisać za pomocą statycznej struktury pojęć, która składa się z „atomów znaczenia"[102].

Chomsky odszedł z czasem od empiryzmu i skłonił się ku racjonalizmowi. Doszedł mianowicie do wniosku, że człowiek wyposażony jest w zdolności pozwalające na przyswajanie wiedzy i dowolne, niezdeterminowane przez środowisko działanie, na które mają jedynie pewien wpływ bodźce zewnętrzne. Gdy mówimy o tych zdolnościach, mówimy według Chomsky'ego o umyśle i powinniśmy skupić się przede wszystkim na wrodzonej zdolności posługiwania się językiem.

Chomsky zwrócił uwagę na uniwersalność pewnych jednostek fonologicznych, syntaktycznych i semantycznych[103]. Nie chodziło mu jednak o konieczność występowania tych, jak je nazywał, *materialnych uniwersaliów* we wszystkich językach, lecz o możliwość ich zdefiniowania i zidentyfikowania za pomocą ogólnoteoretycznych definicji. Obok uniwersaliów materialnych, Chomsky wskazał również na istnienie uniwersaliów *formalnych*, czyli ogólnych zasad wyznaczających formę reguł gramatycznych poszczególnych języków i sposób operowania tymi regułami.

Według Chomsky'ego, wszystkie ludzkie języki są strukturalnie do siebie podobne a fakt ten można tłumaczyć na wiele sposobów, powołując się choćby na wspólne cechy fizjologiczne i psychologiczne użytkowników każdego języka. Jednak za najwłaściwsze wytłumaczenie Chomsky uznał genetyczne wyposażenie człowieka w tak zwaną *zdolność* językową. Założenie podobieństwa strukturalnego wszystkich ludzkich języków, opartego na wrodzonej zdolności językowej, wiąże się z wprowadzeniem pojęcia gramatyki uniwersalnej – ogólnej dedukcyjnej teorii struktury języka, obejmującej wszystkie możliwe systemy komunikacji. Konstruowanie takiej teorii uważał Chomsky za główne zadanie lingwistyki, którego realizacja miała doprowadzić do ustalenia uniwersalnych i istotnych własności ludzkiego języka.

Według Chomsky'ego, człowiek rodzi się z wiedzą o zasadach gramatyki uniwersalnej i predyspozycją do korzystania z tych zasad na podstawie zasłyszanych wypowiedzi. Zasady te, zdeterminowane przez strukturę lub sposób działania mózgu, stanowią składnik tego, co

[99] J. Bobryk, *op. cit.*, s. 40.
[100] Była to formalna analiza pojęć przypisanych wybranym słowom.
[101] Można tu zauważyć wpływ neonatywizmu Chomsky'ego.
[102] …czyli „konstelacji cech lub składników semantycznych" – J. Bobryk, *op. cit.*, s. 42.
[103] Widać tu wpływ poglądów rosyjskiego językoznawcy, Romana Jakobsona.

nazywamy umysłem. Chomsky pojmował umysł jako z góry zaprogramowaną przez naturę, „wyizolowaną od świata i oddzieloną od innych umysłów”[104] monadę, w której odbywają się procesy poznawcze[105].

Idealna kompetencja językowa stanowiła dla Chomsky’ego nie tylko formalny opis języka, ale była także dla niego empirycznym bytem natury neurofizjologicznej. Pogląd ten wpłynął na empiryczny charakter twierdzeń wypowiadanych przez przedstawicieli współczesnej psychologii kognitywnej, którzy przedstawiają formalny model struktur poznawczych w postaci programu komputerowego i uznają, że jego *realizacja* przez maszynę cyfrową pozwala symulować procesy poznawcze. Założenie identyczności przebiegu tych procesów w maszynie i w ludzkim umyśle wynika z pomieszania twierdzeń opisujących rzeczywistość empiryczną z twierdzeniami dotyczącymi modelu formalnego. Jak zauważa Jerzy Bobryk, traktowanie modelu formalnego jako modelu „określonej części rzeczywistości empirycznej” wymaga od jego twórców „empirycznej weryfikacji zgodności modelu z rzeczywistością” i zobowiązuje do „pełnej świadomości tego, które twierdzenia mają charakter opisu rzeczywistości empirycznej, które zaś dotyczą modelu formalnego”[106].

Z jednej strony, mentalizm Chomsky’ego wskazuje na tradycję kartezjańską, według której zachowanie człowieka *nie jest* w pełni zdeterminowane przez bodźce zewnętrzne, czy stany fizjologiczne. Z drugiej strony, odwołania do genetycznego wyposażenia człowieka, struktury i sposobu działania mózgu, pozwalają uznać Chomsky’ego za redukcjonistę, który problem umysłu i ciała uznawał za bezprzedmiotowy ze względu na fakt, iż dzięki rozwojowi nauki „pojęcie «fizyczny» ulegało (…) rozszerzeniu, mającemu na celu objecie wszystkiego, co rozumiemy”[107]. Chomsky nie negował możliwości wytłumaczenia „zjawisk mentalnych” w kategoriach „procesów fizjologicznych i (…) fizycznych, które już znamy”[108].

Prace Chomsky’ego były inspiracją dla poszukiwań takiego rozwiązania problemu relacji między umysłem a ciałem, które byłoby odległe zarówno od dualizmu, jak i redukcjonistycznego behawioryzmu. Rozwiązanie takie stanowiła propozycja „przejścia od substancjalnego do funkcjonalnego ujęcia umysłu”[109].

[104] J. Bobryk, *op. cit.*, s. 53.
[105] Metafora umysłu jako zaprogramowanej przez naturę monady widoczna jest w pracach Jerry’ego A. Fodora, który w ramach reprezentacjonizmu wysnuł koncepcję tzw. języka *mentaleskiego*. Według Fodora, rozpoznanie przedmiotu percepcji poprzedzone jest jego skonfrontowaniem z reprezentacją mentalną i „odbywa się dzięki uruchomieniu odpowiedniego «podprogramu» wkodowanego w naszym mózgu”. Jak zauważyła Urszula Żegleń, „Widać tu wyraźnie wykorzystanie analogii z systemem sztucznym, jakim jest komputer. W ten sposób reprezentacyjna teoria umysłu Fodora (zainfekowana ideami Chomsky’ego) wpisuje się w podejście funkcjonalistyczno-komputacyjne do umysłu” – U. Żegleń, *Filozofia Umysłu*, Adam Marszałek, Toruń 2003, s. 170.
[106] J. Bobryk, *op. cit.*, s. 52.
[107] Cytat za: J. Lyons, *op. cit.*, s. 141.
[108] *Ibidem*, s. 141.
[109] J. Bobryk, *op. cit.*, s. 42.

1.5.3. Koncepcje Alana Turinga

1.5.3.1. Maszyna Turinga

Jedna z najważniejszych współczesnych koncepcji umysłu oparta została na pojęciu *maszyny Turinga*. Jest to koncepcja funkcjonalistyczno-komputacyjna, która dała początek licznym dyskusjom, prowadzącym do różnych odpowiedzi na pytania związane ze współcześnie ujętym problemem relacji umysłu i ciała.

Pomysł maszyny Turinga narodził się z potrzeby znalezienia sposobu stwierdzania prawdziwości zdań dotyczących nieskończenie wielu przypadków za pomocą skończonego zbioru reguł. Alanowi Turingowi (1912-1954) chodziło na przykład o metodę obliczenia liczby rzeczywistej ∏. W artykule „O liczbach obliczalnych, z zastosowaniem do Entscheidungsproblem”[110], Turing odniósł się do nakreślonego przez Dawida Hilberta problemu rozstrzygalności[111], przeformułowując go i nazywając „problemem stopu”[112]. Turing opisał w tym artykule działanie umysłu człowieka dokonującego obliczeń, a następnie stwierdził, że można by było skonstruować maszynę, która działając tak, jak ludzki umysł, wykonywałaby pracę obliczeniową. Turing przyrównał liczącego człowieka do maszyny, nie podejmując z początku tematu zdolności umysłu do działań innych, niż proces obliczania.

Idąc śladami Turinga, wyobraźmy sobie maszynę, przez którą przesuwa się nieskończonej długości taśma podzielona na pola, z których każde może zawierać jeden symbol. Maszyna może w danym momencie odczytywać tylko jedno pole, rejestrując to, co na nim widnieje. Aby odczytać więcej symboli, musi dokonać kolejnych, następujących po sobie odczytów. Maszynie takiej przysługuje skończona liczba stanów, czyli konfiguracji, i może korzystać ze skończonego zestawu symboli. Gdyby było inaczej, niektóre stany albo symbole byłyby dowolnie do siebie zbliżone, a więc nierozróżnialne. Uniemożliwiałoby to pracę maszyny, gdyż jej działanie jest w pełni uzależnione od stanu, w jakim się znajduje i aktualnie odczytywanego symbolu. Wiedząc, że na taśmie może się znaleźć w dowolnym momencie tylko skończona liczba symboli, załóżmy, że maszyna może wykonać tylko następujące czynności: wymazać lub wydrukować symbol, przejść do innego stanu lub pozostać w dotychczasowym albo przesunąć się o jedno pole w lewo lub w prawo.

„Lista instrukcji”, odnosząca się do wszystkich możliwych stanów i symboli, określa jednoznacznie, co maszyna ma zrobić, gdy znajdzie się w danym stanie lub odczyta określony symbol. List takich może być wiele i każda z nich tworzy inną maszynę Turinga[113].

---

[110] „(…) Problemu rozstrzygalności”.

[111] Problem rozstrzygalności to problem odkrycia algorytmu rozstrzygającego, czy dowolna, prawidłowo zbudowana, formuła systemu logicznego jest twierdzeniem tego systemu. Jeśli istnieje skończona procedura pozwalająca w określonym czasie stwierdzić, czy dany element należy do konkretnego zbioru elementów, to zbiór ten jest rozstrzygalny, choć może on być nieskończony.

[112] Problem stopu jest opisany w dalszej części niniejszego rozdziału.

[113] Przy założeniu notacji binarnej, przykład „listy instrukcji” może wyglądać następująco:

1. $S_0 0 = S_0 0 R$
   Jeśli maszyna jest w stanie S_0 i odczytuje symbol 0, to ma pozostać w stanie S_0, pozostawić symbol 0, a następnie przesunąć się o jedno pole w prawo.
2. $S_0 1 = S_1 1 R$
   Jeśli maszyna jest w stanie S_0 i odczytuje symbol 1, to ma przejść do stanu S_1, pozostawić symbol 1, a następnie przesunąć się o jedno pole w prawo.
3. $S_1 0 = S_0 1 STOP$
   Jeśli maszyna jest w stanie S_1 i odczytuje symbol 0, to ma przejść do stanu S_0, wymazać symbol 0 i w jego miejsce wydrukować symbol 1, a następnie zakończyć pracę.

Liczby obliczalne Turing definiuje jako te nieskończone ułamki dziesiętne, które mogą być wydrukowane przez maszynę Turinga wyposażoną na początku w czystą taśmę. Ułamkiem takim jest np. liczba ∏. Lista instrukcji każdej maszyny Turinga jest skończona, a zatem można wszystkie możliwe maszyny Turinga ułożyć w alfabetycznym porządku ich list instrukcji. Zbiór wszystkich możliwych maszyn Turinga jest więc skończony i równoliczny ze zbiorem liczb naturalnych, czyli przeliczalny. Jest to dowód na to, że zbiór liczb obliczalnych, wygenerowanych przez maszynę także jest przeliczalny[114]. Należy uznać, że niewymierna liczba ∏ jest liczbą obliczalną tak samo, jak każda liczba rzeczywista, określona zwykłymi metodami matematycznymi. Zbiór liczb rzeczywistych jest nieprzeliczalny, a więc tylko niektóre spośród nieskończonego zbioru liczb rzeczywistych mogą być obliczalne i będą to liczby wygenerowane przez maszynę Turinga.

W ten sposób, na drodze czysto mechanicznego postępowania, Turing wykazał istnienie liczb nieobliczalnych. Co więcej, chciał dowieść, że za pomocą jego maszyny można wskazać konkretną liczbę nieobliczalną. W tym celu należy wyeliminować ze zbioru maszyn Turinga te, które wskutek wpadania w pętlę nie są zdolne do wygenerowania nieskończonej liczby cyfr. Pozostałe w zbiorze maszyny mogą generować ciągi cyfr będące liczbami obliczalnymi (takimi, jak np. ∏). Każdą liczbę, której jedna z cyfr różni się od cyfr z ciągów generowanych przez maszyny ze zbioru, należy uznać za nieobliczalną.

Pojawia się tu jednak problem nieobliczalności operacji identyfikujących maszyny wpadające w pętlę. Nie może istnieć maszyna Turinga, która na podstawie odczytu listy instrukcji innej maszyny mogłaby rozstrzygnąć, czy wygeneruje ona nieskończony ciąg cyfr. Maszyna taka będzie miała poważny problem z zakończeniem pracy jakąkolwiek odpowiedzią i stąd określenie „problem stopu". Gdyby taka maszyna istniała, można by ją zastosować do niej samej, a to prowadziłoby do sprzeczności.
Turing wykazał, że można za pomocą mechanicznych obliczeń udowodnić istnienie liczb nieobliczalnych, co skłoniło go do przyjęcia tezy, że „funkcja jest efektywnie obliczalna, jeżeli jej wartość można ustalić na drodze czysto mechanicznego postępowania"[115]. Wynika z tego, że o obliczalności funkcji decyduje możliwość ustalenia jej wartości za pomocą maszyny Turinga, która działa na podstawie skończonego zbioru operacji ujętych w liście instrukcji. Również Alonzo Church „udowodnił, że każda efektywnie obliczalna funkcja jest rekursywnie obliczalna"[116], inaczej mówiąc: każda funkcja obliczalna w drodze procedury mechanicznej jest obliczalna z wykorzystaniem skończonego zbioru operacji.

Turing nie poprzestał jednak na teoretycznym rozwiązaniu problemu rozstrzygalności za pomocą abstrakcyjnej maszyny matematycznej. Uznał, że można tak zaprojektować maszynę,

---

4. $S_1 1 = S_1 1 R$
   Jeśli maszyna jest w stanie S_1 i odczytuje symbol 1, to ma pozostać w stanie S_1, pozostawić symbol 1, a następnie przesunąć się o jedno pole w prawo.

Maszyna Turinga działająca na podstawie powyższej listy instrukcji „dodaje jedynkę do dowolnej liczby zapisanej w systemie jedynkowym" – na podstawie: R. Penrose, *Nowy umysł cesarza*, tłum. P. Amsterdamski, PWN, Warszawa 2000, s. 58.

[114] Zbiór nieprzeliczalny to zbiór nieskończony, który nie jest równoliczny ze zbiorem liczb naturalnych, czyli liczb całkowitych nieujemnych. Zbiorami nieprzeliczalnymi są m.in. zbiór liczb rzeczywistych i zbiór liczb niewymiernych. Do zbioru liczb rzeczywistych zaliczają się liczby niewymierne (takie, których nie można przedstawić w postaci ilorazu dwóch liczb całkowitych) i wymierne (takie, które można przedstawić w postaci ilorazu dwóch liczb całkowitych).

[115] A. M. Turing, *Systems of logic based on ordinals*, Proc. Lond. Math. Soc. Ser. 2, 45 (1939), s. 161-228, za: A. Hodges, *Turing*, Amber, Warszawa 1998, s. 27.

[116] U. Żegleń, *Filozofia umysłu*, Adam Marszałek, Toruń 2003, s. 39-40.

aby odczytała każdą listę instrukcji i wykonała pracę każdej innej maszyny. Maszynę taką określił mianem *maszyny uniwersalnej*. Maszyn Turinga, tak jak na przykład egzemplarzy książek, może być potencjalnie nieskończenie wiele, a wśród nich mogą się znaleźć maszyny uniwersalne. Według Turinga, praktyczna maszyna uniwersalna to fizyczna realizacja idei działania umysłu rachmistrza, czyli maszyna o konstrukcji pozwalającej na pracę w oparciu o modyfikowalny i dający się przechować program. Bardzo ważne jest to, że programy i dane mogą być przechowywane w formie symbolicznej. Dzisiejsze programy komputerowe można uznać za maszyny Turinga, a komputer za maszynę uniwersalną. Było to więc pierwsze sformułowanie zasady działania współczesnego komputera i jednocześnie odważna próba modelowania działania ludzkiego umysłu za pomocą fizycznej maszyny, co okaże się szczególnie istotne dla dalszych rozważań.

Turing zauważył, że stany niektórych maszyn „różnią się między sobą na tyle wyraźnie, że uniemożliwia to ich pomylenie", a przejście „od stanu do stanu ma charakter momentalny, skokowy"[117]. Nazwał te maszyny *maszynami o stanach nieciągłych*. Podklasą maszyn o stanach nieciągłych są komputery, które jako maszyny uniwersalne, potrafią imitować działanie dowolnej maszyny o stanach nieciągłych.

Mózg jest częścią układu nerwowego, o którym Turing napisał jednoznacznie, że „niewątpliwie *nie jest* maszyną o stanach nieciągłych", gdyż maszyny takie tak naprawdę „wcale nie istnieją" i „w rzeczywistości wszystko odbywa się w sposób ciągły. Tym niemniej wiele rodzajów rzeczywistych maszyn dobrze jest *uważać za* maszyny o stanach nieciągłych" i „choć istnieją stany pośrednie, to w większości przypadków można je ignorować"[118].

Jak napisał Hodges: „Turing twierdzi, że jedyne cechy mózgu istotne dla myślenia czy inteligencji to te, które ujawnia opis mózgu jako maszyny o stanach dyskretnych", co może oznaczać, że według Turinga mózg jest maszyną o stanach ciągłych, którą „dobrze jest *uważać za*" maszynę o stanach nieciągłych. Jeśli uznalibyśmy, że taki opis mózgu jest wystarczający dla odzwierciedlenia wszystkich cech umysłu, to będziemy musieli zgodzić się z sugestią, że umysł jest maszyną obliczeniową.

Dla Turinga fizyczne różnice między człowiekiem a maszyną nie miały żadnego znaczenia. Jak to dobitnie wyraził: „nie interesuje nas fakt, że mózg ma konsystencję zimnej owsianki" i „jeśli chcemy znaleźć istotne podobieństwa między mózgiem a komputerem, powinniśmy szukać raczej matematycznych analogii funkcjonalnych" [119], a nie analizować ich fizyczną budowę.

Właśnie takimi opiniami, podpartymi pomysłem uniwersalnej maszyny matematycznej, Turing tworzył podwaliny nowej postawy w filozofii umysłu, nazywanej funkcjonalizmem komputerowym, opartym na funkcjonalizmie ortodoksyjnym Hilarego Putnama[120].

[117] A. M. Turing, *Maszyna licząca a inteligencja*, w: *Filozofia umysłu*, pod red. B. Chwedeńczuka, Aletheia, Warszawa 1995, s. 277-278.
[118] *Ibidem*, s. 278.
[119] A. Hodges, *op. cit.*, s. 58.
[120] Zob. podrozdział 2.2.1.

1.5.3.2. Test Turinga

Turing zamierzał zbadać granice obliczalności za pomocą swojej maszyny uniwersalnej i uznał, że możliwe jest urzeczywistnienie tej maszyny w postaci urządzenia elektronicznego, czyli sztucznego mózgu. Możemy tu już mówić o koncepcji sztucznej inteligencji, ponieważ, jak twierdził Turing, „operacje obliczalne obejmują swym zasięgiem zachowania inteligentne”[121]. Uważał ponadto, że nieobliczalność w ogóle nie jest istotna dla inteligencji, czego dowodem miały być sformułowane na podstawie „gry w udawanie” zasady testu Turinga. Zamiarem Turinga było zdefiniowanie inteligencji w taki sposób, aby definicja ta mogła odnosić się nie tylko do ludzi, ale także do maszyn, a właściwie wszystkiego, co można by na jej podstawie uznać za inteligentne. Definicja inteligencji sformułowana przez Turinga była definicją funkcjonalną, opartą na teście przygotowanym specjalnie dla komputerów cyfrowych.

Pytanie „czy maszyny mogą myśleć?” Turing zastąpił pytaniem „czy komputery cyfrowe mogą dobrze wypaść w grze w udawanie?”. Komputer cyfrowy to według Turinga uniwersalna maszyna o stanach nieciągłych, składająca się z rejestru, jednostki wykonawczej i zespołu sterowania. Rejestr służy do przechowywania informacji stanowiących pamięć komputera i zawartych w listach instrukcji, czyli programach. Dzięki zespołowi sterowania, instrukcje te są wykonywane przez jednostkę wykonawczą w odpowiedniej kolejności.

W „grze w udawanie” uczestniczą dwie osoby i komputer cyfrowy. Jedna z tych osób, nazwijmy ją za Turingiem *Prowadzącym*, ma za zadanie rozstrzygnąć wyłącznie na podstawie rozmowy, czy ma do czynienia z komputerem, czy z człowiekiem[122]. Z reguł „gry” wynika, że komputer może udzielać nieprawdziwych odpowiedzi na pytania Prowadzącego. Zadaniem trzeciego uczestnika gry jest niesienie pomocy Prowadzącemu w formie szczerych odpowiedzi na zadane przez niego pytania.

Turing stwierdza, że odpowiednio zaprogramowany i szybki komputer cyfrowy z wystarczająco pojemnym rejestrem może z powodzeniem uczestniczyć w „grze”, ponieważ „bogactwo zachowań maszyny” jest wprost proporcjonalne do wielkości rejestru.

Ważniejsze jest jednak to, że Turing zwraca uwagę na kwestię moim zdaniem fundamentalną dla całej problematyki sztucznej inteligencji. Zastanawia się mianowicie, czy „maszyny dokonują czegoś, co należałoby uważać za myślenie, ale jest dalece odmienne od myślenia człowieka?”[123]. Pytanie to zostaje przez Turinga uznane za „silny zarzut”, którym jednak przejmować się nie trzeba, jeśli jakakolwiek maszyna z powodzeniem jest w stanie brać udział w „grze w udawanie”.

Uważam, że jeżeli w ogóle jest możliwe myślenie nie będące myśleniem ludzkim, to będzie ono prawdopodobnie wymagało wprowadzenia nowych definicji myślenia. Pojęcie myślenia wypracował człowiek. Może ono ewoluować tak, jak na przykład pojęcie duszy. Wśród mnogości definicji myślenia pojawiło się już określenie „myślenia według Turinga”, oparte na obliczeniowej koncepcji umysłu. Dla naszych potrzeb powinniśmy jednak uznać, że gdy mówimy o sztucznej inteligencji, to mamy na myśli inteligencję opartą na myśleniu takim, jak ludzkie. Łatwo bowiem stwierdzić, że maszyna może myśleć, jeśli ograniczymy myślenie

[121] A. Hodges, *op. cit.*, s. 50.
[122] Współcześnie najdogodniejszym sposobem realizacji takiej „gry” byłaby sieć komputerowa, gdyż uczestniczący w niej rozmówcy nie mogą mieć ze sobą bezpośredniego kontaktu.
[123] A. M. Turing, *Maszyna licząca a inteligencja*, w: *op. cit.*, B. Chwedeńczuk (red.), s. 273.

tylko do czynności umysłowych związanych z przetwarzaniem symboli, pamięcią, czy uczeniem się. Dla osób tak drastycznie ograniczających pojęcie myślenia, zdanie testu Turinga przez maszynę byłoby przełomowym wydarzeniem, potwierdzającym pojawienie się sztucznej inteligencji[124]. Nie bez znaczenia pozostaje fakt, że test ten dotyczy właściwie tylko sfery językowej, co pozwala go kojarzyć z teorią lingwistyczną Chomsky'ego i zachowaniami językowymi opisywanymi przez behawiorystów logicznych. Test Turinga obciążony jest w związku z tym całym bagażem krytyki kierowanej przeciw wyżej wymienionym koncepcjom.

Część II
Ideologia Sztucznej Inteligencji

Rozdział II
Ideologiczne podstawy Sztucznej Inteligencji

2.1. Fizykalizm nieredukcjonistyczny i dualizm własności

Po odrzuceniu dualizmu substancji, problem umysłu i ciała formułuje się współcześnie w pytaniu o związek *własności* mentalnych i fizycznych. Jeśli uzna się, że jest to związek przyczynowo-skutkowy, to problematyczne stają się kwestie, co jest przyczyną, a co skutkiem i czy stany mentalne mogą być przyczyną stanów fizycznych, a jeśli tak, to jak to się dzieje? Aby uniknąć takich pytań, należy zaprzestać rozpatrywania związków własności umysłowych i fizycznych w kategoriach przyczynowych. Relacją nie-przyczynową jest relacja *superweniencji*, czyli nadbudowania, więc próbuje się dziś rozwiązać problem umysłu i ciała poprzez odwołanie się do pojęcia superweniencji własności mentalnych na własnościach fizycznych.

Relacja przyczyny i skutku jest asymetryczna i tę cechę ma również relacja superweniencji. Aby coś superweniowało na czymś innym, muszą być spełnione dwa podstawowe warunki:
dwa układy, identyczne pod względem fizycznym, muszą być identyczne pod względem własności superwenientnych (jeśli dwa układy znajdują się w takim samym stanie fizycznym, to znajdują się w takim samym stanie umysłowym); są wprawdzie możliwe zmiany na poziomie fizycznym bez zmian na poziomie mentalnym, ale nie odwrotnie (stany umysłowe nie mogą ulegać zmianom bez zmiany stanów fizycznych).

Dużą rolę w sformułowaniu powyższych warunków odegrały takie dwudziestowieczne koncepcje, jak teoria emergencji i monizm anomalny. Samą teorię superweniencji otwarcie lub „po cichu" akceptują funkcjonaliści (według Jaegwona Kima, funkcjonalizm wręcz *zakłada* superweniencję własności).

[124] Podaje się dziś wiele przykładów maszyn, które zdały test Turinga, co świadczy właściwie tylko o tym, że *zachowanie* tych maszyn zostało zinterpretowane jako inteligentne. Równie dobrze można stwierdzić, że telewizor jest inteligentny, bo *wie*, że gdy naciskam przycisk z numerem jeden, ma przełączyć się na odbiór programu pierwszego.

2.1.1. Teoria emergencji

Redukcjonistyczne i mechanicystyczne podejście badawcze behawioryzmu, panującego w psychologii na początku XX wieku, doczekało się krytyki między innymi ze strony zwolenników teorii *emergencji*[125]. Emergentyści twierdzili, że wynikiem procesu ewolucji jest powstawanie coraz bardziej złożonych poziomów: z czasoprzestrzeni *emerguje* materia, z procesów materialnych wyłania się życie, z procesów witalnych rodzi się umysł i wreszcie z procesów umysłowych rodzi się boskość. Na każdym z tych poziomów „wyłaniają się" (emergują) nowe właściwości (emergenty), których nie da się przewidzieć w oparciu o prawa obowiązujące na poziomach niższych[126].

Za początek świata emergentyści uznawali pierwotną organizację „punkto-chwil", materializujących się jako elektrony, które stanowią podstawę „wyłonienia się" atomów o cechach chemicznych. Podstawą do „wyłonienia się" życia jest poziom fizyczny. Odpowiednio wysoki stopień zintegrowania materii żywej skutkuje emergencją właściwości stanów mentalnych.

Było to pierwsze sformułowanie wielowarstwowego modelu świata, który zastąpił kartezjański model rzeczywistości składającej się z dwu, ontologicznie niezależnych od siebie dziedzin. „Wyłaniające się" właściwości nazywane były zamiennie właściwościami *emergentnymi* lub właściwościami *superwenientnymi*. Zgodnie z tym ujęciem, cechy umysłowe są nadbudowane na mereologicznie skonfigurowanych cechach fizycznych i w ten sposób od nich zależne, choć do nich nieredukowalne[127].

Według teorii emergencji, stopień złożoności niektórych obiektów fizycznych pozwala scharakteryzować te obiekty tylko i wyłącznie poprzez odwołanie się do całkiem nowych właściwości, których nie ma jeszcze na fizycznym poziomie elementarnym i których nie można identyfikować z właściwościami tego poziomu. Emergentyści uważali, że te nowe, wyłaniające się właściwości, nie mają charakteru czysto pojęciowego, lecz są obserwowalne i dają się stwierdzić empirycznie na drodze eksperymentów.

Emergentyzm bronił się przed zarzutem epifenomenalizmu sfery mentalnej tezą, że właściwości wyższego poziomu (właściwości emergentne) oddziałują przyczynowo na obiekty zaliczane do niższego poziomu. Ten rodzaj przyczynowości znany jest pod nazwą *downward causation*. Emergentyzm, przypisując własnościom mentalnym wewnętrzne, niefizyczne moce przyczynowe, okazał się formą dualizmu. Nie był to już jednak dualizm substancjalny, lecz *dualizm własności*.

Emergentyzm to teoria materialistyczna, a materializm przyjmuje istnienie tylko jednej substancji. Głosząc w sprawie substancji tezę monistyczną, nie może być mowy o dualizmie substancjalnym, co jednak rodzi pytanie, jak mają się do siebie własności tej jednej substancji. Jeśli stwierdzi się, że wszystkie własności można sprowadzić na przykład do własności fizycznych, to jest to redukcjonizm. Jeśli dodatkowo uzna się, że w rzeczywistości mamy do czynienia tylko z własnościami fizycznymi, to jest to teza monistyczna w sprawie własności. Materialistyczny monizm własności opiera się jednak na założeniu, że „w pewnym ważnym sensie «fizyczne» implikuje «niementalne»" (i odwrotnie), co w rezultacie jest

[125] Jednym z najbardziej znanych *emergentystów* był filozof australijski Samuel Alexander (1859-1936).
[126] Angielskie słowo *emergence* pochodzi od łacińskiego *emergo* – wynurzyć, wydobyć.
[127] Mereologia to teoria relacji części do całości. W przypadku teorii emergencji chodzi o to, że cechy całości „wyłaniają się" z relacji i własności charakteryzujących części tej nowej całości.

kolejną odmianą dualizmu, nazwanego przez Johna R. Searle'a „dualizmem pojęciowym", charakterystycznym także dla dualizmu substancjalnego. Materializm jest więc według Searle'a „subtelnym wykwitem dualizmu"[128].

Teoria emergencji była krytykowana za nieprecyzyjność i niejasność używanych w niej terminów[129], prowadzącą czasami do stwierdzenia nadnaturalnego charakteru „wyłaniania się" umysłu. Była to jednak pierwsza teoria reprezentatywna dla ważnego we współczesnej filozofii umysłu materializmu nieredukcjonistycznego, którego kontynuacją stały się takie koncepcje, jak mocny funkcjonalizm (w odróżnieniu od redukcjonistycznego słabego funkcjonalizmu), czy teoria superweniencji.

2.1.2. Monizm anomalny i teoria superweniencji

Jak już ukazałem, w sprawie relacji umysłu i ciała, termin „superweniencja" został po raz pierwszy użyty przez emergentystów. Na początku lat siedemdziesiątych wykorzystał go również Donald Davidson, formułując swoją teorię monizmu anomalnego[130].

Davidson uważał, że pojęcia odnoszące się do zjawisk umysłowych nie dają się w żaden sposób zredukować do pojęć fizykalnych[131]. Davidson oparł monizm anomalny na trzech przesłankach. Założył mianowicie, że: po pierwsze, pomiędzy niektórymi zdarzeniami mentalnymi a zdarzeniami fizycznymi zachodzi interakcja przyczynowa; po drugie: przyczynowość ma charakter nomologiczny, więc wszystkie zdarzenia, które oddziałują przyczynowo z innymi zdarzeniami, podlegają takim samym ścisłym deterministycznym prawom, jakim podlegają zdarzenia fizyczne; po trzecie: nie istnieją żadne deterministyczne prawa psychologiczne, ani psychofizyczne.

Ostatnia przesłanka wskazuje, że psychologia (tak samo, jak biologia) nie jest nauką ścisłą, ponieważ zjawiska umysłowe nie dają się objąć prawami deterministycznymi, są *anomalne*. Pozorną niezgodność pomiędzy ostatnią, a pozostałymi przesłankami, Davidson tłumaczył twierdząc, że mimo, iż zdarzeń mentalnych nie da się analitycznie zredukować do zdarzeń fizycznych (zgodnie z postulowanym w ostatniej przesłance anomalizmem zjawisk umysłowych), to zdarzenia mentalne są jednak zależne od fizycznych jako zdarzenia na nich nadbudowane, czyli superwenientne[132].

Davidson sprzeciwiał się behawiorystycznym tezom opartym na teorii identyczności rodzaju (*type-type*). Swój sprzeciw tłumaczył nieprecyzyjnością intuicyjnego pojęcia „zdarzeń mentalnych". Davidson uznał, że podstawą przy opisie wszystkiego, co uznaje się za

[128] J. R. Searle, *op. cit.*, 1999, s. 47-48.
[129] Chodzi na przykład o termin „ewolucja", czy niezdecydowanie poszczególnych emergentystów co do tego, na jakie poziomy należy dzielić rzeczywistość.
[130] Wcześniej termin „superweniencja" (łac. *supervenio* – iść na czymś) pojawił się w *Etyce Nikomachejskiej* Arystotelesa i rozprawach św. Tomasza z Akwinu. Leibniz używał tego terminu do formułowania tez filozoficznych, a George E. Moore wykorzystał go w swej etyce. Do użycia pojęcia „superweniencji" zainspirował Davidsona ten ostatni filozof. Obecnie termin ten jest wykorzystywany w odniesieniu do fizyki, chemii, biologii, psychologii i innych dziedzin, z nadzieją zrozumienia wszystkich tych poziomów opisu rzeczywistości bez konieczności redukowania jednych do drugich.
[131] Wyraźnie zarysowuje się tutaj wspomniany wcześniej dualizm pojęciowy, który Davidson godził z ontologicznym monizmem.
[132] Własności superwenientne mają u Davidsona charakter czysto pojęciowy, co odróżnia jego teorię od emergentyzmu, który postuluje możliwość empirycznego badania tych własności.

umysłowe, jest odwołanie się do intencjonalności, która cechuje każdą postawę orzekającą (*propozycjonalną*)[133]. Postawy orzekające wyrażane są za pomocą zdań zawierających *czasowniki mentalne*[134]. Tego typu zdania stanowią, według Davidsona, wystarczająco precyzyjny opis zdarzeń fizycznych, które na jego podstawie uznaje się za „zdarzenia mentalne".

Superweniencję zdarzeń mentalnych na zdarzeniach fizycznych Davidson rozumiał „w ten sposób, iż nie mogą istnieć dwa zdarzenia identyczne pod wszystkimi względami fizycznymi, ale różniące się pod pewnym względem mentalnym, lub też tak, że przedmiot nie może zmienić się pod pewnym względem mentalnym nie zmieniając się zarazem pod względem fizycznym"[135]. W zdaniu tym zawarte są dwie definicje *superweniencji*, przy czym pierwsza odnosi się do zdarzeń, a druga do obiektów. Podsumowaniem ich może być stwierdzenie, że dwa zdarzenia lub obiekty nie różnią się pod względem psychologicznym, jeśli są nierozróżnialne pod względem fizykalnym. Nie ma bowiem takich zdarzeń i obiektów, które są nierozróżnialne pod względem ich własności podstawowych (subwenientnych), różniąc się przy tym pod względem własności nadbudowanych (superwenientnych). Jeśli każda zmiana pod względem fizycznym pociąga za sobą zmianę pod względem mentalnym, to zmianę na poziomie umysłu można odnieść do odpowiadających im procesów zachodzących w mózgu[136].

W późnych latach siedemdziesiątych teoria superweniencji oddzieliła się od monizmu anomalnego, niosąc ze sobą założenia, które można uznać za bardzo istotne dla współczesnej filozofii umysłu. Istotne jest, że zgodnie z tą teorią, własności mentalnych nie można redukować do własności fizycznych, czy uznawać ich za własności fizyczne, ponieważ cechy nadbudowane nie mogą być cechami podstawowymi. Odpowiadało to tezom funkcjonalistycznym, zgodnie z którymi własności mentalne mogą być realizowane przez własności fizyczne, ale są do nich niesprowadzalne.

Jak zauważa Jaegwon Kim, sama „teoria superweniencji umysłu i ciała nie daje nam teorii relacji, w jakiej umysł pozostaje do ciała"[137], gdyż nie udziela odpowiedzi na pytanie, *dlaczego* w odniesieniu do tego, co mentalne i fizyczne, powinna zachodzić relacja superweniencji. Emergentyści uważali superweniencję umysłu i ciała za coś oczywistego, czego wyjaśniać nie trzeba. Według Kima, wyjaśnienie takie dał dopiero funkcjonalizm, który własności mentalne ukazał jako własności drugiego rzędu, zdefiniowane na „pierwszorzędnych" własnościach fizycznych.

2.2. Funkcjonalizm

Aby zrozumieć, skąd wzięła się współczesna idea konstruowania „myślących maszyn" i dostrzec związane z nią dylematy, należy odnieść się do leżącego u jej podstaw funkcjonalizmu, czyli teorii, która nakazuje charakteryzować stan umysłu w kategoriach roli funkcjonalnej tego stanu i dopuszcza jego dowolną realizację.

[133] Nie można mieć nadziei, czy wierzyć bez żadnego kontekstu intencjonalnego. Można tylko mieć nadzieję na *coś* lub wierzyć w *coś*.
[134] Czasownikami mentalnymi są na przykład takie słowa, jak *wierzyć*, *życzyć*, *lubić*, itp.
[135] Cytat za: J. Kim, *op. cit.*, s. 14.
[136] Stany mentalne można z tego punktu widzenia traktować jako „cienie" stanów fizycznych, co pokazuje, że teoria superweniencji nie broni się przed zarzutem epifenomenalizmu.
[137] *Ibidem*, s. 21.

W połowie XX wieku, w reakcji na słabości behawioryzmu psychologicznego, Putnam sformułował teorię, dla której inspiracją była koncepcja systemu algorytmicznego, zwanego „maszyną Turinga" oraz sporządzony przez Turinga opis testu mającego dowieść, że dana maszyna myśli. Zgodnie z funkcjonalizmem w sformułowaniu Putnama, każdą istotę posiadającą umysł można uważać za maszynę Turinga (o skończonej ilości stanów), której operacje mogą zostać określone przez program. Wynika z tego, że myślenie zależy od odpowiedniej *organizacji funkcjonalnej*, czyli programu. Pozwoliło to Putnamowi analizować stosunek pomiędzy mózgiem a umysłem, jako stosunek pomiędzy komputerem a programem i postrzegać umysł jako oprogramowanie mózgu.

Funkcjonalizm stanów maszynowych Putnama bywa współcześnie określany mianem funkcjonalizmu skrajnego lub ortodoksyjnego, będącego źródłem dla tzw. funkcjonalizmu komputerowego. Uznaje się przy tym, że przedmiotem badań funkcjonalizmu ortodoksyjnego były wszystkie procesy poznawcze, natomiast funkcjonalizm komputerowy skupia się wyłącznie na badaniu ludzkiej i maszynowej inteligencji, używając przy tym aparatury pojęciowej charakterystycznej dla teorii Hilarego Putnama.

2.2.1. Funkcjonalizm Hilarego Putnama

W roku 1960 Hilary Putnam opublikował artykuł *Minds and Machines*, w którym wyłożył podstawowe założenia teorii, którą sam nazwał „funkcjonalizmem". W tym właśnie artykule po raz pierwszy nadany został filozoficzny sens terminowi „realizacja", który na mocy zarysowanej przez Putnama analogii komputerowej, wszedł do powszechnego użycia w omawianiu kwestii dotyczących relacji umysłu i ciała. Tezy przedstawione w artykułach Putnama, opublikowanych w latach sześćdziesiątych i siedemdziesiątych, zostały później uznane przez samego ich autora za zbyt skrajne pod względem ontologicznym, co sprawiło, że stanowisko przez te tezy reprezentowane zwykło się nazywać funkcjonalizmem *ortodoksyjnym*. Jednak to właśnie ten rodzaj funkcjonalizmu, mocno wsparty behawioryzmem, najbardziej przyczynił się do sformułowania twierdzeń stanowiących podstawę ideologii Sztucznej Inteligencji.

Jak zauważa Kim, funkcjonalizm „traktuje własności mentalne jako własności funkcjonalne, określone za pomocą ich ról – jako przyczynowe pośredniki pomiędzy sensorycznymi wejściami i zachowaniowymi wyjściami"[138]. W *mocnej* wersji funkcjonalizmu panuje zupełna dowolność w sprawie materii, w której realizują się pośredniki, natomiast redukcjonistyczny *słaby* funkcjonalizm dopuszcza ich wieloraką realizację, ale tylko w nośnikach fizycznych[139]. Funkcjonalizm ortodoksyjny Putnama jest mocnym funkcjonalizmem, choć nie był nim od samego początku.

W artykule *Minds and Machines* Putnam przedstawia „logiczny opis" przykładowej maszyny Turinga w formie tabeli odpowiadającej „liście instrukcji"[140] oraz opis przedstawiający ją

[138] *Ibidem*, s. 28.

[139] Reprezentantem słabego funkcjonalizmu jest Jaegwon Kim i David Armstrong. Jest to teoria redukcjonistyczna, w której sferę mentalną funkcjonalnie redukuje się do sfery fizycznej, co wymaga opisanej przez Kima *funkcjonalizacji* tego, co umysłowe.

[140] Maszynę Turinga i jej opis w postaci „listy instrukcji" przedstawiłem w poprzednim rozdziale. „Lista instrukcji" bywa także nazywana „tabelą zachowań", natomiast Putnam używa dla jej określenia terminu *machine table*.

jako mogący znaleźć się w skończonej ilości stanów automat, wyposażony w taśmę oraz urządzenia drukujące i „skanujące". Zwraca przy tym uwagę, że „«logiczny opis» maszyny Turinga nie zawiera żadnej specyfikacji *fizycznej natury* tych «stanów» – a tak naprawdę, fizycznej natury całej maszyny. (…) «Maszyna Turinga» jest *abstrakcyjną* maszyną, która może zostać fizycznie zrealizowana na prawie nieskończoną ilość różnych sposobów"[141] (Putnam ograniczał tu jeszcze możliwość realizacji maszyny Turinga do nośników fizycznych).

W dalszym ciągu rozważań, Putnam odróżnia „stany logiczne" od „stanów strukturalnych" maszyny, powołując się na różnicę w punktach widzenia logika i inżyniera, która polega na tym, że ten ostatni traktuje awarię jakiejś części maszyny jako jeden z jej stanów. Putnam rozważa następnie możliwość odczytania tego stanu przez samą maszynę i stwierdza, że po wyposażeniu jej w „organy sensoryczne", które powodowałyby wydruk na taśmie odpowiednich symboli, maszyna mogłaby na podstawie ich odczytu identyfikować własne stany[142]. Konkluzja jest według Putnama następująca: „maszyna, która jest zdolna zidentyfikować przynajmniej niektóre z własnych stanów strukturalnych, znajduje się w położeniu analogicznym do człowieka, który może wykryć, ze zmiennym stopniem pewności, niektóre, choć nie wszystkie, wady w funkcjonowaniu swego ciała"[143].

Jeśli przyjmie się, że człowiek jest jakimś rodzajem maszyny Turinga, to jest to maszyna przetwarzająca bodźce docierające do organów sensorycznych, traktowane jako dane *wejściowe*, na odpowiednie zachowania *wyjściowe*. Jeśli ktoś znałby aktualny *stan* „ludzkiej maszyny", to na podstawie danych zawartych na „liście instrukcji", mógłby stwierdzić, jaki będzie jej kolejny *stan* i jak się ona zachowa.

Warto zauważyć, że dla behawioryzmu stany mentalne były tylko wrodzonymi dyspozycjami do zachowania, które nie odgrywają żadnej przyczynowej roli[144]. W funkcjonalizmie stany mentalne są stanami funkcjonalnymi, scharakteryzowanymi przez ich role przyczynowe.

Putnam stwierdza, że zachowanie maszyny Turinga można opisać na dwóch poziomach: logicznym (jako „listę instrukcji") albo fizykalnym (jako strukturalny projekt inżyniera albo fizyka), a następnie zauważa, że takie same poziomy opisu odnoszą się do ludzkiej psychologii. Fizykalny poziom opisu reprezentuje behawioryzm, który „dąży do dostarczenia kompletnego fizykalistycznego opisu ludzkiego zachowania w terminach odnoszących się do chemii i fizyki"[145].

Według Putnama, ludzkie procesy umysłowe można opisać na bardziej abstrakcyjnym poziomie, zupełnie pomijając kwestię ich realizacji. Opis taki odpowiadałby logicznemu opisowi maszyny Turinga i zawierałby prawa określające kolejność następujących po sobie „stanów mentalnych" oraz relacje do „zwerbalizowanych myśli". „Doznania" pełniłyby funkcję symboli drukowanych na taśmie. Tak uzyskany opis mógłby, zdaniem Putnama, wspomóc psychologię, której braki są dotkliwie odczuwalne nie ze względów

[141] H. Putnam, *Minds and Machines*, w: H. Putnam, *op. cit.*, 1975, s. 371.
[142] Putnam zauważa przy tym, że człowiek (tak, jak maszyna bez „organów sensorycznych") nie zawsze może stwierdzić, w jakiej kondycji jest jego wyrostek robaczkowy.
[143] *Ibidem*, s. 372.
[144] Rolę przyczynową pełnią w behawioryzmie bodźce, a własności całego organizmu mogą tylko umożliwiać reakcję na te bodźce.
[145] *Ibidem*.

metodologicznych, lecz dlatego, że „stany mentalne i «doznania» ludzi nie tworzą tak przyczynowo domkniętego systemu, jak «konfiguracje» maszyny Turinga”[146, 147].

Tak, jak funkcjonalna organizacja maszyny może zostać opisana w odniesieniu do kolejności występowania stanów logicznych, tak też można dokonać opisu człowieka, odnosząc się do kolejności występowania stanów mentalnych i werbalizacji, pomijając w każdym z tych opisów sposób „fizycznej realizacji” tych stanów[148]. I tak, jak można opisać maszynę, na przykład w odniesieniu do jej własności fizycznych, tak też można opisać istotę ludzką. Według Putnama, poziom „logiczny” i poziom „fizykalny” to dwa niesprowadzalne do siebie poziomy opisu. Tak, jak w przypadku konkretnej maszyny „bycie w stanie A” i „włączenie przełącznika X” to dwa różne stany, tak samo w przypadku człowieka „stan bólu” i „pobudzenie włókien nerwowych C_4” są różnymi stanami, dotyczącymi odpowiednio własności logicznych i fizycznych.

Idąc śladem sugerowanej przez Putnama analogii komputerowej, dochodzi się do wniosku, że to, iż „lista instrukcji”, czyli program maszyny (*software*) może zostać zrealizowany w różnych fizycznych nośnikach (*hardware*), dotyczy również procesów umysłowych. Teza o wielorakiej realizacji wynika z założenia, że poziom umysłowy jest poziomem logicznym i że myślenie opiera się wyłącznie na obliczeniach.

Obliczeniowa teoria umysłu stanowiła spadek po wynikach rozmyślań Alana Turinga[149]. Putnam przejął od Turinga tezę, że opis ludzkiego umysłu, jako maszyny Turinga o stanach dyskretnych (nieciągłych), jest opisem zupełnie wystarczającym dla uchwycenia własności istotnych dla myślenia. Potwierdzeniem tej tezy miało być założenie przez Putnama możliwości opisu procesu umysłowego, jako skończonego szeregu występujących po sobie stanów mentalnych. Według Putnama, sporządzenie takiego opisu może zostać dokonane bez przywiązywania wagi do sposobu fizycznej realizacji procesów umysłowych. Przy takich przesłankach umysł sprowadza się do *funkcji* systemu o dowolnej budowie fizycznej.

W artykule z 1973 roku, zatytułowanym *Philosophy and our mental life*, Putnam umacnia swój funkcjonalizm, wprost wyrażając twierdzenie o pełnej autonomiczności sfery mentalnej i wyciągając z niego daleko idące wnioski. Stara się wykazać, że jedynym sposobem rozwiązania problemu umysłu i ciała jest teoria oparta na pojęciu *izomorfizmu funkcjonalnego*.

Według definicji Putnama, „dwa systemy są funkcjonalnie izomorficzne, jeżeli *stany jednego systemu korespondują ze stanami innego systemu z zachowaniem relacji funkcjonalnych*”[150, 151]. Wynika z tego, że jeśli mielibyśmy poprawną teorię działania jakiegoś systemu X (np.

[146] *Ibidem*, s. 373.

[147] Jak zauważa Putnam w eseju *Umysł a ciało*, już w XVII wieku uświadomiono sobie, że świat fizyczny jest układem przyczynowo domkniętym na sposób wyrażony językiem fizyki Newtona, której twierdzenia łatwo można ująć liczbowo. Problem wydaje się tkwić w przyczynowym domknięciu świata *nie-fizycznego*, o ile przyjmie się jego istnienie i o ile chce się do tego świata stosować prawa przyczynowe.

[148] Pod organizację funkcjonalną podpada według Putnama myślenie, czy rozwiązywanie problemów.

[149] Jak już zauważyłem w pierwszym rozdziale, za spadkodawców obliczeniowej teorii umysłu należy uznać także Hobbesa i Locke’a.

[150] H. Putnam, *Philosophy and our mental life*, w: H. Putnam, *op. cit.*, 1975, s. 292.

[151] Putnam dokładnie wyjaśnia, co ma na myśli, odwołując się do przykładu dwóch sekwencyjnie działających maszyn Turinga. Jeżeli w jednej z tych maszyn stan A występuje po stanie B wtedy i tylko wtedy, gdy w drugiej maszynie stan A’ występuje po stanie B’, to mamy do czynienia z funkcjonalnym izomorfizmem tych stanów. Aby uważać obie maszyny za funkcjonalnie izomorficzne, muszą być zachowane relacje funkcjonalne zawarte w

„program psychologii"), to system ten byłby funkcjonalnie izomorficzny z systemem Y, jeśli możliwe by było takie odwzorowanie wszystkich własności i relacji zdefiniowanych w systemie Y, że wszystkie odniesienia do systemu X zostałyby zastąpione przez odniesienia do systemu Y oraz wszystkie własności i symbole relacji zawarte w tej teorii dałyby się reinterpretować na podstawie tego odwzorowania.

Według Putnama, brak „programu psychologii" nie przeszkadza w ukazywaniu różnic między jakąkolwiek możliwą psychologiczną teorią człowieka a jego opisem fizykalnym. Putnam stwierdza, że „dwa systemy o zupełnie odmiennej budowie mogą być funkcjonalnie izomorficzne", co oznacza, że „komputer wykonany z elektrycznych komponentów może być [funkcjonalnie, G.B.] izomorficzny z tym zrobionym z zębatek i trybików albo z ludzkimi urzędnikami używającymi papieru i ołówka"[152]. Opisy fizykalne różniłyby się, ale opisy funkcjonalne byłyby takie same.

W eseju *Umysł a ciało* Putnam zaznacza, że „różnice między robotem a człowiekiem (w każdym razie, jeśli chodzi o organizację funkcjonalną) można sprowadzić do drobnych szczegółów fizycznych i chemicznych. (…) dlaczego miałoby się nie powiedzieć, że ów robot jest osobą, z której mózgiem tak się złożyło, iż ma on w sobie więcej metalu, a mniej wodoru i węgla?"[153]. W dalszym ciągu Putnam sugeruje, że taki robot mógłby być świadomy nawet wtedy, gdyby składał się z miniaturowych ludzi przekazujących sobie znaki, funkcjonalnie zorganizowanych tak samo, jak ludzkie neurony[154].

Jak widać, teza o wielorakiej realizacji doprowadziła Putnama do wniosku, że opisywanie stanu mentalnego poprzez omawianie jego fizycznej czy chemicznej realizacji jest absurdalne. „Ojciec funkcjonalizmu" podważył podstawowe materialistyczne założenia teorii identyczności i behawioryzmu stwierdzając, że stany umysłowe nie mogą być identyczne ze stanami fizycznymi[155].

Funkcjonalizm Putnama nabiera mocy, gdy autor przedstawia koncepcję dwóch „światów równoległych". W jednym z tych światów ludzie mają mózgi, a w drugim – kartezjańskie dusze. Putnam zwraca uwagę, że jeśli te mózgi i dusze są funkcjonalnie izomorficzne, to liczy się tylko ich funkcjonalna struktura, a nie materia, czy substancja. Według Putnama, zarówno Kartezjusz, jak i materialiści, kierowali się założeniem, że jeśli człowiek składa się tylko z

---

„listach instrukcji" tych maszyn. Tak więc, jeżeli widniejące na taśmie Π powoduje przejście jednej maszyny do stanu A tak samo, jak drugiej do stanu A', to maszyny te są funkcjonalnie izomorficzne.

[152] *Ibidem*, s. 292-293.

[153] H. Putnam, *Umysł a ciało*, w: *op. cit.*, B. Chwedeńczuk (red.), s. 206.

[154] Według Putnama, nie ma powodu sądzić, że jeśli „miniaturowi ludzie" są świadomi, to cały robot musi być mniej świadomy od nich. Cytując Putnama: „jesteśmy w pewnym sensie społecznością małych zwierząt. Nasze komórki są w pewnym sensie pojedynczymi zwierzętami. Kto wie, może mają one trochę czucia? Poza i ponad *naszym* czuciem" – *ibidem*, s. 207. Tok myślenia Putnama przypomina tutaj rozważania Daniela Dennetta, który stwierdza, że „*składamy się* z robotów (…). Zatem coś, co składa się z robotów, *może* przejawiać autentyczną świadomość" i dodaje później: „jest tak, *jak gdyby* te komórki i zbiorowiska komórek były malutkimi, prostymi *podmiotami*, wyspecjalizowanymi urzędnikami, racjonalnie realizującymi swoje obsesyjne zadania". Dennett powołuje się tu na pewną strategię interpretacji zachowania bytów, którą nazywa *nastawieniem intencjonalnym* i która polega na traktowaniu każdego bytu tak, *jak gdyby* był podmiotem racjonalnym. Można dojść do wniosku, że Putnam przyjął w swych rozważaniach bardzo podobną strategię (cytaty pochodzą z: D. C. Dennett, *Natura umysłów*, tłum. W. Turopolski, CIS, Warszawa 1997, s. 36 i 39).

[155] Warto tu wspomnieć, że (zgodnie ze sformułowaną przez Leibniza zasadą identyczności tego, co nierozróżnialne), jeśli dwie rzeczy są nierozróżnialne, to znaczy, że są jedną i tą samą rzeczą. Jeżeli więc wyróżni się stany mentalne i stany fizyczne, to nie może być mowy o ich identyczności, bo gdyby były identyczne, to byłyby tymi samymi stanami, co podważałoby zasadność powyższego wyróżnienia.

materii, to ludzkie zachowanie i umysł można wyjaśnić fizykalnie[156]. Aby dowieść, że było to założenie niesłuszne, Putnam zauważa, że wyjaśnianie powinno zawsze odbywać się na poziomie, który jest adekwatny metodologicznie, czyli najodpowiedniejszy do osiągnięcia określonych celów i dosięgający istotnych dla danego problemu praw[157].

Według Putnama, procesy umysłowe zachodzą na poziomie autonomicznym, wymagającym odrębnego opisu. Putnam nie odchodzi od koncepcji, że człowiek jest maszyną Turinga. Dochodzi jednak do wniosku, że opis umysłu nie powinien odpowiadać logicznemu opisowi takiej maszyny, gdyż stany psychologiczne nie są prostymi stanami maszyny Turinga[158]. Opis taki wymagałby przyjęcia stanów, w których człowiek mógłby znaleźć się w każdej chwili i które, w pełni wyszczególniając aktualną kondycję człowieka i uwzględniając wpływający na pamięć proces uczenia się, *natychmiast* powodowałyby przejście do wielu innych stanów. Założenia takie byłyby konieczne, ponieważ maszyna Turinga może być w określonej chwili tylko w jednym ze stanów[159], a uczenie się nie jest w jej przypadku nabywaniem nowych stanów, lecz tylko przyswajaniem informacji wydrukowanej na taśmie. Człowiek, znajdujący się w jakimś konkretnym stanie maszynowym, może jednocześnie odczuwać ból, wydawać i słyszeć jęk, itd. Putnam twierdzi, że teoria psychologiczna nie może odnosić różnych stanów psychologicznych do jednego stanu fizycznego, gdyż opis taki nie wyjaśniałby żadnych istotnych dla psychologii kwestii.

Putnam konkluduje, że wpływ koncepcji maszyn Turinga na filozofię umysłu jest znaczący, choć nie we wszystkich punktach pozytywny. Koncepcji tej można zawdzięczać wprowadzenie pojęcia organizacji funkcjonalnej, które wiąże się z odróżnieniem abstrakcyjnej struktury od jej konkretnej realizacji i twierdzeniem, że ta sama abstrakcyjna struktura może być realizowana na wiele różnych sposobów. Jednak wyjaśnienie tego pojęcia na przykładzie maszyn Turinga, jako systemów o bardzo ograniczonej i specyficznej organizacji funkcjonalnej, może kusić do przyjęcia założenia, że podobnie prostą organizację funkcjonalną posiada człowiek. Założenie to stało się ideologiczną podstawą dla rozpoczęcia badań nad sztuczna inteligencją.

Leżąca u podstaw tych badań obliczeniowa teoria umysłu doczekała się ostrej krytyki[160]. Sam Hilary Putnam w końcu od niej odszedł. W eseju *Wiele twarzy realizmu* (1987) filozof ten pisze, iż jego „«Funkcjonalizm» głosił, że istoty myślące są *ustrojowo plastyczne*”[161].

---

[156] Kartezjusz obawiał się takiej możliwości, a materialiści uważali ją za coś oczywistego.

[157] Jeśli kula nie mieści się w kwadratowym otworze, to nie wyjaśnia się tego molekularną strukturą tych figur, ani nie wylicza się możliwych trajektorii ich ruchu. Najprostszym, i przez to najlepszym wyjaśnieniem, jest odniesienie się do ogólnych zasad geometrii.

[158] Inne zdanie Putnam wyrażał w *Minds and Machines*, gdyż - jak twierdzi – był wtedy „pod zbyt wielkim wpływem wizji redukcjonistycznej”. Następuje tu zdecydowane umocnienie koncepcji funkcjonalistycznej.

[159] Dlatego każdy możliwy stan musiałby spełniać wszystkie wymienione warunki.

[160] Najbardziej znany argument przeciw obliczeniowej teorii umysłu opiera się na twierdzeniu Gödela, które mówi, że „dla dowolnego systemu formalnego *M*, zawierającego pewną część arytmetyki liczb naturalnych, można skonstruować w języku systemu *M* takie zdanie, którego nie da się udowodnić w *M*, ani też nie da się udowodnić w *M* negacji tego zdania” – za: *Encyklopedia filozofii* pod red. T. Hondericha, t. I, tłum. J. Łoziński, Zysk i S-ka, Poznań 1998, s. 297. Analiza tego argumentu wymagałaby odrębnej rozprawy, wobec czego ograniczę się jedynie do stwierdzenia, że według Gödela umysł ludzki nie jest maszyną Turinga, gdyż w przypadku człowieka często można mieć do czynienia nie tylko z mechanicznymi procedurami wnioskowania, ale także z *intuicją*. Umysł ludzki nie podlega więc ograniczeniom, którym podlega system formalny. Już Alan Turing przewidział ten argument, doszukując się sposobu na jego odparcie w koncepcji maszyn uczących się w drodze działania losowego, czego wzorem była dla Turinga teoria ewolucji.

[161] H. Putnam, *Wiele twarzy realizmu*, w: H. Putnam, *Wiele twarzy realizmu i inne eseje*, tłum. A. Grobler, PWN, Warszawa 1998, s. 338.

Następnie stwierdza, że porzucił teorię głoszącą izomorfizm struktury funkcjonalnej człowieka i struktury funkcjonalnej maszyny Turinga, ponieważ „stany umysłu są nie tylko ustrojowo plastyczne, ale też *obliczeniowo plastyczne*”[162]. Wynika z tego, że możliwe jest istnienie organizmów, które myślą dokładnie to samo, różniąc się nie tylko pod względem fizycznym, ale również pod względem „oprogramowania”.

2.2.2. Jerry A. Fodor – język „mentaleski”

Jak zauważa Putnam, „na gruncie modelu mózgu jako systemu poznawczego przypominającego komputer, mózg ma język, ma jakiś język wewnętrzny (może on być wrodzony, a może być mieszanką «języka» wrodzonego, czyli reprezentacji, *oraz* języka publicznego). Niektórzy filozofowie wymyślili nawet nazwę dla tego hipotetycznego języka mózgu – jest to język «mentaleski»”[163]. Putnam ma tu na myśli Jerry’ego A. Fodora, uważanego za jednego z czołowych przedstawicieli słabej odmiany funkcjonalizmu, która wsparta osiągnięciami psychologii kognitywnej, bywa nazywana psychofunkcjonalizmem[164]. Jedną z głównych inspiracji dla tego amerykańskiego filozofa był ortodoksyjny funkcjonalizm Putnama. Drugim ważnym czynnikiem, mającym wpływ na formułowane przez Fodora twierdzenia, była głoszona przez Chomsky’ego koncepcja wrodzonych struktur gramatycznych.

Według reprezentacyjnej teorii umysłu, której źródła można odnaleźć w filozofii Leibniza, Locke’a i Hume’a, umysł jest systemem będącym reprezentacją świata. Obliczeniowa teoria umysłu głosi, że system ten stanowi złożoną strukturę procesów obliczeniowych, których przebieg jest zależny nie tylko od informacji percepcyjnej, ale także od wewnętrznych właściwości umysłu[165]. Fodor proponuje połączyć reprezentacyjną teorię umysłu z „metafora komputerową”.

Sugestia taka zawarta została między innymi w eseju o żartobliwym tytule: *Jak grać w reprezentacje umysłowe – poradnik Fodora*. Fodor stwierdza w nim, że „komputery pokazują nam, w jaki sposób powiązać własności semantyczne *symboli* z własnościami przyczynowymi”[166] i dodaje, że powiązanie takie można w nich uzyskać poprzez *syntaktykę* symbolu, jeśli jego strukturę syntaktyczną uzna się za abstrakcyjną własność *kształtu* tego symbolu. W mniemaniu Fodora, kształt ten potencjalnie determinuje rolę przyczynową symbolu, co oznacza, że syntaktykę symbolu należy uważać za jedną z jego własności fizycznych drugiego rzędu.

Jak twierdzi Fodor, „syntaktyka symbolu może determinować przyczyny i skutki jego egzemplarzy w taki mniej więcej sposób, jak geometria klucza determinuje, jaki zamek można nim otworzyć”[167]. W dalszym toku rozważań Fodor odwołuje się do logiki formalnej i zauważa, że implikacja, jako relacja syntaktyczna między symbolami, może naśladować relację semantyczną, jeśli sąd reprezentowany przez jeden symbol implikuje sąd

[162] *Ibidem*, s. 339.
[163] H. Putnam, *Umysł a ciało*, w: *op. cit.*, B. Chwedeńczuk (red.), s. 195-196.
[164] Na podstawie: *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/.
[165] Jeśli uzna się umysł za monadę, to w tym przypadku okazuje się, że jednak ma ona okna. Według Fodora organizm podlega wpływom środowiska. Natywizm tego filozofa opiera się na twierdzeniu, że pojęcia nie mogą być nabywane tylko z zewnątrz, że są one wrodzone (internalizm).
[166] J. A. Fodor, *Jak grać w reprezentacje umysłowe – poradnik Fodora*, tłum. A. Putko, w: *Modele umysłu*, pod red. Z. Chlewińskiego, PWN, Warszawa 1999, s. 40.
[167] *Ibidem*, s. 41.

reprezentowany przez inny symbol. Środowiskiem, w którym „przyczynowa rola egzemplarza symbolu musi odpowiadać roli inferencyjnej sądu” [168, 169] przez ten symbol wyrażanego, mógłby być według Fodora komputer, który reagowałby wyłącznie na syntaktyczne własności symboli i którego działanie polegałoby tylko na przekształcaniu jednego symbolu w drugi, co następowałby pod warunkiem, że między tymi symbolami zachodzi jakaś relacja semantyczna (na przykład taka, jak między przesłanką a wnioskiem w dedukcji).

Fodor konkluduje, że komputer można z takiej perspektywy uznać za „pośrednika” między przyczynowymi a semantycznymi własnościami symboli. Można to wyjaśnić w poniższy sposób.

1. Fizyczna własność symbolu, czyli jego kształt (symbole mogłyby mieć na przykład formę *p* i *q*) jest własnością pierwszego rzędu;
2. syntaktyczna własność symbolu, jako własność drugiego rzędu, jest specyfikowana w kategoriach przyczynowej roli odpowiedniego kształtu symbolu, która może być określona w „liście instrukcji” maszyny, czyli przez program komputera. Oto bardzo uproszczony przykład takiej „listy instrukcji”:

   S_0 *p* P *q* R S_1
   S_0 *q* P *p* R S_2
   S_1 *p* P *q* STOP
   S_1 *q* P *q* STOP
   S_2 *q* P *p* STOP
   S_2 *p* P *p* STOP[170]

3. Semantyczna własność symbolu, jako własność drugiego rzędu, jest specyfikowana w kategoriach inferencyjnej roli sądu odpowiadającego danemu symbolowi.

Maszyna Turinga, działająca według powyższej „listy instrukcji”, mogłaby być jedną z szeregu działających maszyn, które odzwierciedlałyby ludzki umysł. Ta akurat maszyna odpowiadałaby wnioskowaniom logicznym, które można zapisać w schemacie *modus ponendo ponens*.

Przypuśćmy, że ktoś wnioskuje o czymś na podstawie swoich dotychczasowych przekonań. Na przykład, na podstawie przekonania, że „świeci słońce” (*p*) oraz, że „jeśli świeci słońce, to jest dzień” ($p \rightarrow q$), nabywa pewności, że „jest dzień” (*q*). *Własność semantyczna* symboli *p* i *q* odpowiada według Fodora ich roli we wnioskowaniu $[(p \rightarrow q) \Lambda p] \rightarrow q$ (można je wyrazić zdaniem: „jeśli słońce świeci, to jest dzień, a słońce świeci, więc jest dzień”).

Zgodnie z twierdzeniami Fodora, między własnościami przyczynowymi symbolu, zdeterminowanymi przez jego własności fizyczne (kształt [tutaj poziom I]), a własnościami semantycznymi [poziom III] pośredniczą własności syntaktyczne [poziom II]. Role przyczynowe symboli konstytuują (na poziomie II) strukturę fizyczną z tych symboli złożoną. Ich role inferencyjne natomiast, konstytuują (na poziomie III) strukturę semantyczną symboli. Fodor twierdzi, że obie te struktury są funkcjonalnie izomorficzne.

[168] Chodzi tutaj o rolę sądu we wnioskowaniu.
[169] *Ibidem*.
[170] Instrukcje te można wyrazić takimi zdaniami, jak: „Jeśli maszyna jest w stanie S_0 i odczytuje z taśmy symbol *p*, to zawsze drukuje w jego miejscu symbol *q*, następnie przesuwa się o jedno pole w prawo i przechodzi do stanu S_1”. STOP oznacza, że maszyna ma zakończyć pracę i zadanie można uznać za wykonane.

Wynika z tego, że „w konstrukcji komputera rola przyczynowa jest sprzężona z treścią dzięki wykorzystaniu odpowiedniości między syntaktyką symbolu a jego semantyką”[171]. Zgodnie z myślą Fodora, sprzężenie takie dotyczy również neuronalnej struktury mózgu i formalnej struktury ludzkiego wnioskowania.

Fodor mówi o treści sądów wyłącznie dla ukazania ich struktury logicznej. Uważa on, że między znaczeniem jakiegoś słowa, a jego funkcją, zachodzi relacja tożsamości; że znaczenie słowa po prostu jest jego funkcją, którą można nazwać funkcją *semantyczną*[172].

Fodor postuluje stworzenie konkurencyjnej dla behawioryzmu teorii opartej na założeniu, że umysł jest rodzajem przedstawionego wyżej komputera. Teoria taka, aby wyjaśnić regularność relacji treściowych między przyczynowo powiązanymi myślami, musiałaby, według Fodora, przyjąć istnienie symboli w postaci reprezentacji umysłowych, posiadających własności semantyczne i syntaktyczne. Jak zauważa Fodor, obraz umysłu jako „maszyny kierowanej przez syntaktykę” jest równoznaczny z twierdzeniem, że „teoria procesów umysłowych może zostać w pełni wyłożona bez jakichkolwiek odniesień do semantycznych własności stanów umysłowych”[173]. Zaznacza jednak przy tym, że jest daleki od zaprzeczenia istnieniu tych własności i zauważa, że choć „syntaktyczna teoria operacji umysłowych obiecuje redukcyjne wyjaśnienie *inteligentnych* mechanizmów myślenia”, to nie stanowi ona „teorii *intencjonalności* myśli” i jeśli jest prawdziwa, to według Fodora „problem intencjonalności stanów umysłowych jest głównie – być może wyłącznie – problemem semantyczności reprezentacji umysłowych”[174], który dopiero wymaga rozwiązania.

W opinii Fodora, trudności w uporaniu się z tym problemem są spowodowane faktem, że założenie, iż procesy umysłowe są przekształceniami poszczególnych reprezentacji umysłowych, pociąga za sobą wymóg, aby wszystkie treści umysłowe były reprezentowane *explicite*. Tymczasem, jeśli przyjmie się istnienie postaw propozycjonalnych komputera, to intencjonalność większości z nich można uważać najwyżej za epifenomen działań, które maszyna wykonuje, kierując się wieloma bardzo szczegółowymi regułami. Fodor zauważa, że takiego „cienia pracy maszyny” nie można ująć *explicite* w jeden symbol umysłowy, a co ważniejsze, maszyny działają tylko tak, *jak gdyby* stosowały się do ustalonych w programie reguł[175], a więc *jak gdyby* intencjonalnie[176].

---

[171] *Ibidem*, s. 42.

[172] Taka teoria znaczenia ma swój wyraz nie tylko w nauce kognitywnej, gdzie bywa nazywana *semantyką proceduralną*, ale również w filozofii, jako *semantyka ról pojęciowych* lub *semantyka przyczynowa*. Semantyka ról pojęciowych wiąże się z *redundancyjną* teorią prawdy, zgodnie z którą prawdziwość sądu zależy od tego, czy sąd ten wchodzi w skład twierdzeń akceptowanych przez osobę wyrażającą ten sąd. Zwolennikiem tej teorii jest między innymi amerykański filozof Hartry H. Field, twórca *funkcjonalistycznej filozofii matematyki*. Koncepcję semantyki proceduralnej skrytykował w 1980 roku Hilary Putnam stwierdzając, że język „mentaleski” nie może opisywać świata zewnętrznego, gdyż „rozumienie” tego języka oparte jest wyłącznie na „mózgowym programie” posługiwania się nim, a przecież – jak twierdzi Putnam – każdy program stosuje się nie do przedmiotów istniejących w świecie zewnętrznym, lecz do tego, co jest zawarte *wewnątrz* komputera.

[173] *Ibidem*, s. 42.

[174] *Ibidem*, s. 48.

[175] Fodor powołuje się przy tym na Daniela Dennetta i nadmienione w jednym z wcześniejszych przypisów nastawienie intencjonalne. Warto tu również wspomnieć fragment eseju *Nauka o poznawaniu* autorstwa Johna Searle’a, który stwierdza w nim, że komputery nie kierują się regułami w tym samym sensie, co ludzie, lecz jedynie działają zgodnie z określonymi procedurami formalnymi. Chodzi tu o użycie metafory - „komputery nie stosują się do reguł, tak jak ludzie, one tylko działają, tak jak gdyby przestrzegały reguł” – J. Searle, *Nauka o poznawaniu*, w: J. Searle, *Umysł, mózg i nauka*, tłum. J. Bobryk, PWN, Warszawa 1995, s. 43.

Przedstawione wyżej połączenie reprezentacyjnej teorii umysłu z metaforą komputerową nie ukazuje jeszcze w pełni koncepcji języka „mentaleskiego", choć dotyka jej podstawowych zasad. Aby zasady te ugruntować, należy odnieść metaforę komputerową do pojęcia stanu umysłowego. Posłużę się w tym celu pracą, która powstała w wyniku współpracy Jerry'ego Fodora i innego amerykańskiego filozofa, Neda Blocka. Wpływ na jej treść miał również Hilary Putnam, co pozwala uznać tę rozprawę za odzwierciedlającą poglądy trzech ważnych postaci, wypowiadających się na temat funkcjonalizmu i zadających sobie pytanie: *Czym nie są stany psychiczne?*[177].

Rozprawa ta jest cenna również dlatego, że zawiera często formułowaną przez przedstawicieli funkcjonalizmu, zdecydowaną krytykę fizykalizmu redukcyjnego, a co za tym idzie, również behawioryzmu, i co najciekawsze – jednej z teorii funkcjonalistycznych, zwanej *Teorią Identyczności Stanów Funkcjonalnych* (*TISF*), która wydała się Fodorowi po prostu „za mało abstrakcyjna".

Za podstawową wadę behawioryzmu Fodor[178] uznaje konsekwencje empiryczne tej teorii, czyli nikłe prawdopodobieństwo spełnienia wymogu takiego dopasowania predykatów behawioralnych do konkretnych typów stanów psychicznych, aby to, czy jakiś organizm znajduje się w określonym stanie, zależało od możliwości orzekania o tym stanie przyporządkowanego mu predykatu behawioralnego. Skompletowanie takiego zbioru par predykatów psychologicznych i behawioralnych Fodor uważa za niewykonalne, gdyż zachowanie (lub dyspozycja do zachowania) stanowi w przypadku człowieka bardzo złożoną funkcję nie tylko stanu wejść sensorycznych, ale także wspomnień, przekonań i pragnień.

Według Fodora, wystarczająco udowodnione jest twierdzenie, że takie same wspomnienia, przekonania i pragnienia mogą mieć organizmy o odmiennych stanach fizycznych, co podważa założenia fizykalizmu redukcjonistycznego, który zakładał identyczność typów stanów psychicznych z typami stanów fizycznych tak samo, jak behawioryzm sugerował identyczność typów stanów psychicznych z odpowiednimi typami zachowań[179]. Fodor stwierdza, że „typy stanów psychicznych nie odpowiadają typom stanów fizycznych"[180], uznając wykorzystanie teorii identyczności typów za przyczynę klęski nie tylko fizykalizmu i behawioryzmu, ale także wyżej wspomnianej *TISF*.

TISF głosi, że „dla każdego organizmu, który w ogóle spełnia predykaty psychologiczne, istnieje jedyny najlepszy *opis*, taki, że każdy stan psychiczny tego organizmu jest tożsamy z

---

[176] Z problemem intencjonalności borykają się nie tylko funkcjonaliści, ale także badacze sztucznej inteligencji. Opisanie poglądów związanych z tym problemem i wniosków, jakie mogą z nich wypływać, wymagałoby odrębnej pracy.

[177] Pytanie to stanowi tytuł rozprawy omawianej w dalszym ciągu tego podrozdziału. Ważne jest, aby pamiętać, że gdy mówi się w niej o stanach psychicznych, to chodzi nie tylko stany organizmu ludzkiego, ale także stany każdego innego organizmu. Gdy natomiast formułowane są twierdzenia o stanach mentalnych, zaznacza się zwykle, że chodzi o stany organizmu posiadającego osobowość.

[178] W dalszym ciągu, jeśli będę się odnosił do rozprawy *Czym nie są stany psychiczne?*, to przywołując nazwisko Jerry'ego Fodora, będę zakładał, że można by było zastąpić je nazwiskami współautorów - Blocka i Putnama.

[179] Fodor powołuje się przy tym na teoretyczne rozważania Putnama, teorię ekwipotencjalności neurologicznej Lashley'a („każda z wielu rozmaitych funkcji psychicznych może być spełniana przez wiele różnych struktur mózgu") oraz wyprowadzoną z darwinowskiej doktryny konwergencji tezę o podobieństwie zachowania różnych gatunków istot żywych – cytat za: J. Fodor, *Czym nie są stany psychiczne?*, tłum. T. Baszniak, w: *op. cit.*, B. Chwedeńczuk (red.), s. 60-61.

[180] *Ibidem*, s. 60.

jednym ze stanów jego urządzenia sterującego, zrelatywizowanych do tego opisu" [181,182]. Warto jednak zauważyć, że *TISF* utożsamia ze sobą nie tylko konkretne stany, ale także zawiera twierdzenie, że „każdy typ stanów psychicznych jest tożsamy z (pewnego typu) stanem urządzenia sterującego"[183], ujętym w jedynym najlepszym opisie organizmu[184].

Według *TISF*, każdy organizm najlepiej jest opisać jako rodzaj „automatu probablistycznego", czyli maszyny Turinga, której stany są wzajemnie powiązane ze sobą, wejściami i wyjściami, za pośrednictwem prawdopodobieństw przepływu danych w urządzeniu sterującym tego organizmu[185]. Opis taki powinien, według *TISF*, spełniać podstawowy wymóg: musi być opisem tych typów stanów organizmu, które jedno-jednoznacznie odpowiadają konkretnym typom stanów psychicznych[186]. Warunku tego nie spełnia jednak *każdy* opis określonego organizmu jako maszyny Turinga, lecz tylko ten, który jest najlepszy z możliwych („jedyny najlepszy"), można by powiedzieć – harmonijny, „współmożliwie" uporządkowany tak, jak świat opisany przez Leibniza[187].

Leibniz formułował jednak twierdzenia ontologiczne, a *TISF* nie jest teorią ontologiczną, ponieważ milczy na temat tego, do jakich systemów można zastosować predykaty psychologiczne. Opisowi zgodnemu z *TISF* mogą bowiem podlegać nie tylko osoby, czy komputery, ale każdy przedmiot materialny, a nawet dusza, co potwierdza, że nie wychodzimy poza granice mocnego funkcjonalizmu, dla którego przyjęcie koncepcji monad nie byłoby niczym dziwnym. Tam bowiem, gdzie identyfikuje się własności umysłowe z własnościami drugiego rzędu, tam mamy do czynienia z *liberalizmem* w sprawie realizacji stanów mentalnych. Uwalnia to *TISF* od trudności związanych z ontologicznym statusem „nośnika" stanów mentalnych, lecz nasuwa wątpliwość, czy *na pewno* status ten nie odgrywa żadnej roli w kwestii umysłu. Problem ten poruszę w ostatnim rozdziale pracy[188]. Tutaj

[181] Urządzenie sterujące maszyny Turinga to urządzenie zawierające „listę instrukcji". Dane wejściowe przechodzą przez urządzenie sterujące, które na podstawie „listy instrukcji" (*machine table*) kieruje działaniem maszyny. Urządzenie sterujące charakteryzuje każdy stan w terminach zbioru instrukcji regulujących zachowanie maszyny, ilekroć znajdzie się ona w tym stanie.

[182] *Ibidem*, s. 65.

[183] *Ibidem.*

[184] *TISF* jest więc teorią identyczności rodzaju (*type-type*).

[185] W deterministycznej („standardowej") maszynie Turinga „lista instrukcji" określa dla każdego stanu maszyny jedno wyjście i jeden kolejny stan. W probabilistycznej maszynie Turinga „lista instrukcji" określa dla każdego stanu maszyny wiele wyjść lub kolejnych stanów, zgodnie z przyporządkowanym im prawdopodobieństwem wystąpienia.

[186] Według Fodora, każdy organizm można opisać jako maszynę Turinga, która obliczając funkcję stanu fizycznego tego organizmu (jeśli jest ona obliczalna), będzie ten stan symulować, nie dając przy tym żadnej gwarancji, że jest on tożsamy z jakimkolwiek stanem psychicznym. Putnam zwrócił uwagę, że wszystko można opisać jako tzw. „zerową maszynę Turinga". Maszyna taka otrzymuje dane na wejściu, ale nie zmienia swego stanu i nie przesyła żadnej informacji na wyjście. Opis takiej maszyny uzyskuje się, pomijając założenia dotyczące tego, co należy traktować jako wyjście organizmu, a co jako zmianę stanu organizmu. To, czy jakiś system realizuje określona maszynę Turinga, jest względne wobec tych założeń. Ta względność odzwierciedla w pewien sposób różne poziomy wyjaśniania, o których pisał Putnam.

[187] Można by sądzić, że jeśli u autora *Monadologii* za harmonię tę odpowiedzialny był Bóg, to zwolennicy *TISF* najprawdopodobniej chwytaliby się tutaj teorii ewolucji.

[188] Sądzę, że właśnie teza o wielorakiej realizacji stanów mentalnych, uznawana za największe osiągnięcie funkcjonalizmu, jest jednym z powodów, dla którego funkcjonalizm zawiódł i nie przedstawia wiarygodnej teorii umysłu. Choć na pewnym poziomie opisu teza o wielorakiej realizacji sprawdza się (np. pułapką na myszy może być przedmiot z drewna, drutu i sprężyny albo pojemnik gazu paraliżującego podłączony do fotokomórki), to w przypadku ludzkiego umysłu zastąpienie materii inną materią wymagałoby także zachowania *wszystkich* funkcji neuronów, a właściwie całej organizacji funkcjonalnej komórek nerwowych mózgu, z całym układem nerwowym i *sensorium*, co byłoby właściwie równoważne stworzeniu żyjącego organizmu, który musiałby dodatkowo wchodzić w interakcje z otoczeniem i uczyć się.

natomiast ważne jest to, że Fodor stara się argumentować za porzuceniem *TISF* na rzecz koncepcji, która byłaby jeszcze bardziej liberalna, gdyby nie pretendowała do miana empirycznej teorii psychologicznej, nacechowanej dla odmiany ludzkim *szowinizmem*[189].

Wspomniana wyżej trudność, z jaką według Fodora nie może poradzić sobie behawioryzm, wiąże się z niemożnością przypisania każdego ze stanów psychicznych jakiemuś konkretnemu zachowaniu. Fodor twierdzi, iż trudność ta wynika z tego, że skutki behawioralne wywołują zwykle dwa rodzaje interakcji, zachodzących między wieloma stanami psychicznymi. Zachowanie może być bowiem wynikiem sekwencyjnych lub równoczesnych interakcji między stanami psychicznymi. Interakcje sekwencyjne dają się opisać w terminach maszyny Turinga, która w jednej chwili może znajdować się tylko w jednym stanie. Odpowiada to dotychczas ukazanym założeniom *TISF*. Przyjęcie istnienia równoczesnych interakcji między stanami psychicznymi wymaga jednak od *TISF* znacznej korekty w koncepcji „jedynego najlepszego" modelu organizmu. Okazuje się mianowicie, że model taki powinien być zbiorem działających równolegle wielu maszyn Turinga. Można wprawdzie, jak zauważa Fodor, algorytmicznie zredukować każdy zbiór maszyn do jednej, ale nie będzie to opis „najlepszy z możliwych". Wynika z tego, że według *TISF*, najlepiej jest opisywać organizm, jako działający równolegle procesor. Jest to spostrzeżenie cenne, ale nie chroniące *TISF* przez dalszą krytyką, której sedno nie tkwi wcale w tym, jaki rodzaj interakcji zachodzi między stanami psychicznymi, lecz w odpowiedzi na pytanie: *Czym nie są stany psychiczne?*

Jednym z argumentów przeciwko *TISF* jest zarzut, że teoria ta nie radzi sobie z problemem własności jakościowych, czyli *qualiów*. Fodor odpiera ten atak stwierdzeniem, że różnice między własnościami jakościowymi stanów psychicznych są nieistotne, jeżeli nie determinują różnic funkcjonalnych. Nie da się jednak uniknąć problemu qualiów podobnie, jak problemu statusu ontologicznego realizatora stanów mentalnych. *TISF* nie znajduje odpowiedzi na argument wynikający z faktu, że teoretycznie mogą istnieć dwa funkcjonalnie identyczne stany psychiczne, z których tylko jeden posiada własność jakościową[190].

---

[189] Funkcjonalizm ujęty jako empiryczna teoria naukowa stanowiąca nurt psychologii, został nazwany przez Neda Blocka psychofunkcjonalizmem. Szowinizm tej teorii oznacza, według Blocka, to samo, co szowinizm fizykalizmu, a mianowicie odmówienie posiadania stanów umysłowych zbyt wielu rzeczom, łącznie z rzeczami, o których intuicyjnie można stwierdzić, że mogą je mieć. Szowinizm fizykalizmu różni się jednak u swych podstaw od szowinizmu psychofunkcjonalistycznego. Pierwszy wynika z redukcji umysłu do ludzkiego układu nerwowego, choć – jak sugeruje Block – intuicja podpowiada, że i zwierzęta mogą posiadać jakieś „zalążki" stanów mentalnych. Drugi szowinizm opiera się na domniemaniu, że empiryczne badania naukowe mogą wykazać, iż istoty równoważne nam pod względem funkcjonalnym, z wyjątkiem mechanizmów poznawczych, mogą być na tyle różne od nas pod względem psychologicznym, że można by było zaprzeczyć, iż istoty te w ogóle myślą. Block dzieli pod tym względem funkcjonalizm na *zdroworozsądkowy* (liberalny) i *psychologiczny* (szowinistyczny). *TISF* opiera się na utożsamieniu ról przyczynowych stanów umysłowych z rolami przyczynowymi stanów bazowych. Nie muszą to być koniecznie stany mózgu, czy nawet stany fizyczne, więc nie jest to teoria z gruntu szowinistyczna, choć może się nią stać w pracach naukowych neurofizjologa. Według słabego funkcjonalizmu własności mentalne mogą zaistnieć tylko jako specyficzne dla danego rodzaju (np. ból człowieka, ból kota) i nie mogą być wielorako realizowane. Słaby funkcjonalizm jest funkcjonalizmem tylko ze względu na wykorzystanie specyfikacji funkcjonalnej. Fizykalistycznej teorii identyczności nie powinno się identyfikować z opisywaną tu *Teorią Identyczności Stanów Funkcjonalnych*, gdyż ta pierwsza może być powiązana z funkcjonalizmem tylko w jego słabej odmianie, która poprzez funkcjonalną specyfikację, epistemologicznie *identyfikuje* tylko stany umysłowe na podstawie ich ról przyczynowych identycznych z rolami przyczynowymi stanów fizycznych (stany umysłowe nie są tu stanami funkcjonalnymi, tylko stanami fizycznymi, poziom stanów umysłowych jest tylko funkcjonalnym poziomem opisu stanów fizycznych z uwagi na ich role przyczynowe). Słaby funkcjonalizm jest więc szowinistyczny poprzez swój związek z fizykalizmem.

[190] Zmuszałoby to na przykład do twierdzenia, że ktoś czuje ból w chwili, gdy nie czuje nic.

W ramach dygresji warto zaznaczyć, że problem qualiów to zmora wszystkich funkcjonalistów, o której wspomnę jeszcze w rozdziale dotyczącym kłopotów, jakie czekają każdego, kto chce oprzeć swoją koncepcję umysłu na założeniach funkcjonalizmu. Według Fodora, założeń tych nie dotyka niżej przedstawiony zarzut, który ma ostatecznie obalić *TISF*.
Funkcjonalizm zakłada, że identyczność typów stanów psychicznych uwarunkowana jest ich wzajemnymi relacjami, oraz ich relacjami z sensorycznymi wejściami i behawioralnymi wyjściami organizmu. Według *TISF* natomiast, identyczność ta zależy wyłącznie od identyczności ze stanami urządzenia sterującego organizmu. Fodor uważa, że błędem jest jedno-jednoznaczne przyporządkowywanie stanom psychicznym ich odpowiedników w postaci poszczególnych stanów urządzeń sterujących organizmu. W ten sposób, poddając w wątpliwość teorię identyczności poszczególnych egzemplarzy stanów psychicznych z ich maszynowymi odpowiednikami, Fodor podważa prawdziwość teorii identyczności typów tych stanów[191]. Uzasadnieniem tego jest dla Fodora fakt, że „lista instrukcji" maszyny Turinga zawiera skończoną ilość stanów, natomiast nomologicznie możliwych typów stanów mentalnych może być nieskończenie wiele, gdyż ich zbiór jest zbiorem produktywnym[192]. Ponadto, model urządzenia sterującego nie pozwala uwzględnić „relacji strukturalnych" zachodzących między stanami mentalnymi[193].

W konsekwencji swych przemyśleń, Fodor wprowadza rozróżnienie między stanami *urządzenia sterującego* maszyny (scharakteryzowanymi w „liście instrukcji" urządzenia sterującego), a stanami *obliczeniowymi* maszyny (scharakteryzowanymi „w terminach jej wejść, wyjść i/lub stanów urządzenia sterującego"[194]).
Według Fodora, „jeśli chcemy myśleć o psychologii organizmów, w której są one przedstawiane jako automaty, to stany psychiczne organizmów wydają się być analogiczne do stanów obliczeniowych automatu, a nie do stanów jego urządzenia sterującego"[195]. Stany obliczeniowe to stany kalkulacji, dowodzenia, wnioskowania i inne, których opis można sprowadzić do opisu wejść, wyjść, czy stanów urządzenia sterującego[196].
Można teraz powrócić do zaproponowanej przez Fodora psychologicznej teorii umysłu, według której przedstawione wyżej mechanizmy obliczeniowe są genetycznie zakodowane w mózgu człowieka. Mechanizmy te stanowią wrodzony potencjał uczenia się nowych pojęć, którym odpowiadają poszczególne słowa. Pojęcia te są zawarte w reprezentacjach umysłowych, które można uznać za *umysłowe symbole* niewerbalnego języka wewnętrznego, który Fodor nazywa językiem „mentaleskim". Język „mentaleski" opisuje świat zewnętrzny, przy czym odniesienie do przedmiotów występujących w tym świecie następuje na mocy związków przyczynowych, zachodzących między tymi przedmiotami, a symbolami wewnętrznego języka myśli[197]. Dla opisania tego języka posłużyła Fodorowi koncepcja gramatyki generatywnej Chomsky'ego. Język „mentaleski" zawiera według Fodora:

[191] Identyczność typów zakłada identyczność egzemplarzy, więc jeśli zaprzeczy się identyczności egzemplarzy, to zaprzeczy się także identyczności typów.
[192] „O ile zbiór stanów urządzenia sterującego maszyny Turinga może zostać, na mocy definicji, wyczerpująco scharakteryzowany przez wyliczenie, o tyle zbiór stanów mentalnych osoby w najlepszym razie można scharakteryzować przez skończoną aksjomatyzację". Chodzi o takie aksjomaty, jak na przykład: „przekonanie, X", „myśl, że X", „wiara, że X", gdzie X może oznaczać cokolwiek z nieskończonego zbioru – *Ibidem*, s. 76.
[193] Fodor podaje tu przykład braku możliwości zdefiniowania dla potrzeb urządzenia sterującego stanu „bycia składową", czyli wyrażenia relacji między stanem mentalnym „bycia przekonanym, że X" a stanem „bycia przekonanym, że X i Y".
[194] *Ibidem*, s. 77.
[195] *Ibidem*, s. 79-80.
[196] Chodzi o predykaty w rodzaju „przeprowadzając obliczenia wymagające ośmiu stanów urządzenia sterującego, korzystając z danych wejściowych, maszyna podała na wyjściu wynik równania matematycznego".
[197] Koncepcję znaczenia reprezentacji mentalnych, odwołującą się do ich związków przyczynowych z przedmiotami znajdującymi się w świecie zewnętrznym, Fodor przedstawił dopiero pod koniec ubiegłego wieku,

1. skończoną bazę prostych *pojęć* (reprezentacji, symboli umysłowych), których własność semantyczna jest zależna od roli inferencyjnej odpowiadających im słów i sądów, należących do dowolnego języka naturalnego;
2. nieskończoną ilość *terminów* złożonych, generowanych na podstawie reguł syntaktycznych, których nie da się sprowadzić do pojedynczych symboli umysłowych;
3. nieskończoną ilość *zdań* złożonych, generowanych na podstawie reguł syntaktycznych z symboli umysłowych i terminów złożonych.

Fodor twierdzi, że na „język mentaleski" składa się także wrodzony słownik predykatów, wystarczający do formowania konstrukcji logicznych każdego możliwego języka naturalnego. Wynika z tego, że bez języka wewnętrznego nie byłoby możliwe opanowanie żadnego języka publicznego. Język „mentaleski" Fodor porównuje do kodu maszynowego, czyli obliczeniowego, wewnętrznego języka komputera.

Język publiczny (naturalny, etniczny) to według Fodora odzwierciedlenie komputerowego języka symbolicznego, umożliwiającego przepływ informacji między maszyną a jej użytkownikiem. Fodor twierdzi, że komputer operuje językiem symbolicznym na wejściach i wyjściach, natomiast w celu dokonywania obliczeń używa języka wewnętrznego; i podobnie człowiek – porozumiewa się z innymi ludźmi za pomocą języka publicznego, natomiast *myśli* w języku „mentaleskim".

Komputacyjna teoria umysłu opiera się na założeniu, że myślenie jest procesem obliczeniowym, co pozwala przyjąć hipotezę, że dokonujący odpowiednio skomplikowanych obliczeń komputer również może myśleć; że język wewnętrzny komputera może być językiem „mentaleskim".

Pytanie, czy „komputerowe myślenie" może zaowocować semantyką, jest poważnym wyzwaniem dla zwolenników komputacyjnej teorii umysłu. Jak zauważył Fodor, teoria ta „może zajmować się wyłącznie syntaktyką, czyli wewnętrzną organizacją, a nie semantyką operacji umysłowych"[198]. Twierdzenia takie stanowią punkt wyjścia dla krytyki ideologii Sztucznej Inteligencji. Ideologia ta opiera się na założeniu, że mózg można porównać do syntaktycznej maszyny, która „napędza" maszynę semantyczną, czyli umysł. Wyjaśnienie podstaw takiego porównania pozwala ukazać, w czym tkwi sedno funkcjonalistycznej teorii umysłu przedstawionej przez Fodora. Sięgając do tych podstaw, należy przede wszystkim zwrócić uwagę na pewne rozróżnienia[199].

---

w reakcji na krytykę postawy skrajnie internalistycznej, którą filozof ten przyjął w początkowym etapie formułowania teorii języka wewnętrznego. Autorem tej krytyki był między innymi Hilary Putnam, wysuwający zarzut braku *odniesienia przedmiotowego* reprezentacji mentalnych, którym przypisane są podstawowe *symbole* języka „mentaleskiego". Jako realista, Putnam mówi o braku *ekstensji* pojęć. Można tę krytykę sformułować inaczej, twierdząc, że według wczesnej myśli Fodora, symbole reprezentacji mentalnych są związane z przedmiotem tylko za pomocą formalnych reguł, których sformułowanie wymaga jednak jakiejś „materii", a tę stanowić mogą jedynie reprezentacje „bezpośrednio" odnoszące się do przedmiotów świata zewnętrznego, czego Fodor nie zauważył. W odpowiedzi na ten zarzut, Fodor dokonuje jakby przekształcenia podstawowych symboli umysłowych w swego rodzaju *oznaki* umysłowe, związane z przedmiotami związkami przyczynowo-skutkowymi. Według Fodora, jeśli na przykład ktoś jest przekonany, że świeci słońce, to przekonanie to pojawia się dlatego, że jest ono, jako symbol (czy raczej oznaka), przypisane odpowiedniej reprezentacji mentalnej, która zostaje „wywołana" spośród nieskończonej ilości reprezentacji wskutek działania określonych sygnałów, zarejestrowanych przez organizm za pośrednictwem sensorycznych wejść. Tak, jak dym jest oznaką ognia, tak przekonanie, że świeci słońce można uznać za oznakę określonych sygnałów napływających z otoczenia.

[198] Nie są to słowa Fodora. Zacytowany tu został fragment z: J. Bobryk, *op. cit.*, s. 47.

[199] Wykorzystam tutaj treść rozprawy Neda Blocka pt. *The Mind as the Software of the Brain*, opublikowanej na stronie internetowej http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/msb.html.

Tak, jak słowo „Gdańsk” nie jest miastem Gdańsk, tak symbol „1” nie jest liczbą 1, gdyż tak, jak wiele różnych nazw może odnosić się do tego samego miasta, tak znaki „1”, „jeden”, czy „I” mogą wskazywać (i wskazują) tą samą liczbę 1. Poza tym, tak, jak jedna nazwa może oznaczać różne rzeczy w różnych językach[200], tak jeden symbol może wskazywać różne liczby w zależności od przyjętej notacji[201]. Pewne jest tylko to, że symbole zawsze coś *symbolizują*, przy czym w tych rozważaniach są to akurat liczby.

Przywołam teraz przykład dwóch abstrakcyjnych maszyn, które można nazwać *sumatorem* i *multiplikatorem*. Funkcją sumatora jest dodawanie, a multiplikatora – mnożenie. Rozróżnienie maszyn w oparciu o odpowiedź na pytanie, *co robią*, jest charakterystyczne dla funkcjonalistycznie ukierunkowanych przedstawicieli nauk kognitywnych. Na tej podstawie można przyjąć, że sumator jest *procesorem pierwotnym*, a multiplikator – *wtórnym*. Rozróżnienia takiego można także dokonać, kierując się odpowiedzią na pytanie, *jak działają* obie te maszyny.
Zadaniem sumatora jest dodawanie do siebie pojedynczych cyfr, graficznie symbolizujących określone liczby. Program sumatora napisany jest w notacji binarnej, dlatego wyniki rozwiązanych przez tę maszynę równań są następujące:

0 + 0 = 0
1 + 0 = 1
0 + 1 = 1
1 + 1 = 10

Pierwsze trzy równania są prawdziwe zarówno w notacji binarnej, jak i dziesiętnej, ale prawdziwość ostatniego wymaga odczytania wyniku wyłącznie w notacji binarnej.

Funkcje sumatora, jako procesora pierwotnego, wykorzystane są w procesorze wtórnym, czyli multiplikatorze, który dokonuje operacji mnożenia poprzez ich „rozkład” na pierwotne operacje dodawania. Teoretycy Sztucznej Inteligencji definiują przedmiot swych badań analogicznie do powyższej eksplikacji mnożenia, „poprzez «rozkład» zdolności inteligentnych do sieci zdolności mniej inteligentnych, ostatecznie opartych na w pełni mechanicznych zdolnościach realizowanych przez procesory pierwotne”[202].

Multiplikator ma taki program, aby mnożył *m* razy *n*, dodając *m* do zera tyle razy, ile wskazuje *n*. Maszyna ma tu trzy rejestry: M, N, i A, w których reprezentowane są, odpowiednio, liczby *m* i *n* oraz odpowiedzi *a*. Na początku pracy maszyny rejestr odpowiedzi A jest pusty (A = 0). Po wprowadzeniu do rejestrów M i N reprezentacji *m* i *n*, maszyna sprawdza, czy rejestr N jest pusty, i jeśli *n* = 0, to uznaje, że N = 0 i kończy pracę. Jeśli rejestr N nie jest pusty, to multiplikator odejmuje 1 od N i dodaje *m* do A, a następnie znowu sprawdza, czy N jest pusty, i tak dalej, aż do zakończenia pracy, której wynikiem jest odpowiedź reprezentowana przez *a*[203].

[200] Na przykład słowo „Police” w języku angielskim jest nazwą organu powołanego do ochrony porządku i bezpieczeństwa publicznego, natomiast w języku polskim może oznaczać miasto niedaleko Szczecina.
[201] Symbol „10” w notacji binarnej wskazuje liczbę 2, a w notacji dziesiętnej liczbę 10.
[202] N. Block, *op. cit.*
[203] Przypuśćmy, że multiplikator ma podać wynik mnożenia 2 x 3. W takim przypadku *m* = 2, natomiast *n* = 3. Jako że rejestr N nie jest pusty (N = 3), maszyna odejmuje 1 od 3 i dodaje 2 do rejestru A. Po tej operacji A = 2, ale nie jest to ostateczna odpowiedź *a*, gdyż N nadal nie jest pusty (po pierwszej operacji odejmowania N = 2). Maszyna ponownie odejmuje 1 od N i dodaje *m* do rejestru A (teraz A = 4). Czynność ta musi być powtórzona

Zasada działania multiplikatora nie jest jednak tak istotna, jak fakt, że maszyna ta używa algorytmu niezależnego od notacji, gdyż *m*, *n* i *a* mogą być zapisane w kodzie dziesiętnym lub binarnym. Na tym poziomie notacja nie ma żadnego wpływu na prawidłowość wyników obliczeń, gdyż maszyna mnoży *symbole*, natomiast *liczby* występują w niej na poziomie pierwotnym i mogą być realizowane na przykład tak, jak w komputerze, jako różne impulsy elektryczne o stałej częstotliwości[204]. Również program multiplikatora może być napisany w dowolnej notacji, dogodnej dla programisty tej maszyny. Ważne, aby na wejściach i wyjściach pojawiały się symbole zrozumiałe dla jej użytkownika.

Symbole mają w tym kontekście dwie funkcje: odwzorowywania symboli innymi symbolami oraz odwzorowywania liczb innymi liczbami. Funkcja pierwsza, zwana przez Blocka funkcją *symboliczną*, dotyczy liczb jako symboli, abstrahując od ich znaczenia. Nie następuje tu nic więcej, jak tylko uznanie czegoś fizycznego w maszynie lub na jej wyjściach za symbole, a jakiegoś innego fizycznego aspektu maszyny za wskazanie, że te symbole są wejściami albo wyjściami. To zupełnie wystarczy, aby maszyna, otrzymując na wejściu jakieś symbole, dawała na wyjściu inne symbole. „Funkcja symboliczna opiera się więc na interpretacji struktury przyczynowej maszyny”[205]. Z funkcją tą mamy do czynienia między innymi wtedy, gdy maszyna (na przykład wyżej opisany sumator) odwzorowuje dwa symbole („1” i „0”) jednym symbolem („10”).

Funkcja odwzorowująca liczby innymi liczbami odnosi się już nie do samych symboli, lecz do ich znaczenia (do tego, jakimi są liczbami). Jest to więc funkcja *semantyczna*, która wymaga wybrania jakiejś dowolnej notacji. Warto przypomnieć, że dowolność ta była dopuszczalna również w przypadku maszyny mnożącej, zwanej multiplikatorem[206].

Zgodnie z podstawowymi założeniami obliczeniowej teorii umysłu, funkcja symboliczna i funkcja semantyczna są izomorficzne. W tym kontekście multiplikator jest maszyną semantyczną, „napędzaną” przez pierwotną w stosunku do niej maszynę syntaktyczną, którą jest sumator. Fizyczne aspekty maszyny są interpretowane jako symbole w taki sposób, że pewne symbole na wejściach zawsze wywołują określone symbole na wyjściach (zgodnie ze swoimi funkcjami symbolicznymi). Regularności te są izomorficzne z relacjami zachodzącymi między istotnymi dla użytkownika maszyny semantycznymi wartościami poszczególnych symboli (na przykład dodawaniem, czy mnożeniem). Izomorfizm funkcjonalny wyjaśnia tutaj, w jaki sposób maszyna manipulująca symbolami dodaje i mnoży liczby.

Jak stwierdza Block, „na najbardziej podstawowym poziomie obliczeniowym komputery są przetwornikami symboli i z tego powodu komputerowy model umysłu często jest opisywany jako wizja umysłu [sprowadzonego do – G.B.] manipulowania symbolami”[207]. W tym właśnie sensie mózg uważany jest za maszynę składniową, która „napędza” maszynę semantyczną,

---

jeszcze raz, bo dopiero wtedy N = 0, co oznacza zakończenie pracy i uzyskanie prawidłowej odpowiedzi. Jeśli bowiem maszyna zakończyła pracę i rejestr A = 6, to reprezentuje on odpowiedź *a* = 6.

[204] Nawiasem mówiąc, częstotliwość impulsów nerwowych jest zmienna, co podważa zasadność przyjmowania analogii komputerowej w jej tradycyjnym sformułowaniu. Są już jednak konstruowane komputery oparte na sieciach neuronowych, które działają pod omawianym względem tak, jak ludzki mózg.

[205] *Ibidem*.

[206] Block zwraca uwagę, że ukazanej tu *funkcji semantycznej* nie należy mylić z tak samo nazywaną funkcją odwzorowywania symboli na liczby, do których te symbole się odnoszą. Funkcję taką należałoby uważać za „pośredniczącą” między funkcją symboliczną a funkcją semantyczną w wyżej opisanym sensie.

[207] *Ibidem*.

jaką jest umysł. Według Fodora, jest to możliwe dzięki symbolicznej strukturze mózgu, która w wyniku procesu ewolucji i uczenia się, jest skorelowana z semantyczną strukturą umysłu. Aby odkryć znajdujące się w mózgu symbole, nie trzeba dokonywać trepanacji czaszki, lecz należy poznać semantyczne relacje miedzy stanami mentalnymi, a następnie zidentyfikować symboliczne aspekty tych stanów.

Zgodnie z koncepcja Fodora, znaczenie symbolu stanowi jego funkcja semantyczna, czego sens można wyrazić na przykładzie poniższego zdania, które mimo występujących w nim braków, jest zrozumiałe, nie tracąc swego znaczenia:

„funkcj* symboliczn* jest czymś, co może ktoś docenić po uk*z*niu, j*k j*wi się on* w zd*niu z*wier*jącym zn*jome słow*, których zn*czenie możemy tylko zg*dyw*ć”[208].

2.2.3. Różne oblicza funkcjonalizmu

Próbując uchwycić pojęcie kartezjańskiej duszy zgodnie z tezami funkcjonalizmu, trzeba by było opisać ją za pomocą pojęcia funkcjonalnej organizacji, jako abstrakcyjną strukturę, która może zostać fizycznie zrealizowana na wiele różnych sposobów. Strukturę tę można by było opisać tylko w drodze dokonania funkcjonalizacji jej własności. Byłoby to podejście charakterystyczne dla słabego funkcjonalizmu.

Można by było również rozumieć duszę bardziej tradycyjnie i uznać ją za jeden z możliwych „nośników” innej abstrakcyjnej struktury, zwanej *umysłem*. Byłby to „nośnik” czysto eteryczny, który również posiadałby własną organizację funkcjonalną. Takie czysto teoretyczne rozważania cechują zwolenników funkcjonalizmu mocnego[209].

Przedstawiciele obu nurtów z pewnością by twierdzili, że pomysł uznania własności psychicznych za własności funkcjonalne jest rozsądnym rozwiązaniem problemów wynikających z traktowania umysłu i ciała jako dwóch substancji (interakcyjny dualizm psychofizyczny) lub dwóch odrębnych sfer własności (materialistyczny dualizm własności). Funkcjonalizm nie chroni się jednak przed swoistym dualizmem własności funkcjonalnych i „niefunkcjonalnych”, cechujących materię realizacji tych pierwszych[210].

Funkcjonalistą może być zarówno zwolennik fizykalizmu, jak i psycho-fizykalnego dualizmu, co zależne jest od odpowiedzi na pytania o naturę stanów umysłowych.

Według mocnej wersji funkcjonalizmu, pomiędzy stanami umysłowymi a stanami fizycznymi zachodzi relacja identyczności typów (zgodnie z *Teorią Identyczności Stanów*

[208] *Ibidem*.

[209] Doskonałym przykładem twierdzeń w duchu mocnego funkcjonalizmu może być następująca wypowiedź Putnama: „myślenie o umysłowości, uczuciowości itd. jakiegoś jestestwa jako o aspektach jego struktury czynnościowej pozwala uznać, że najrozmaitsze logicznie możliwe «układy» czy jestestwa mogą w dokładnie tym samym czasie być świadome, przejawiać umysłowość i uczucia itd., a zarazem różnić się pod względem budowy materialnej (a nawet nie składać się z żadnej «materii» w sensie cząstek elementarnych czy pól elektromagnetycznych). Jestestwa o bardzo rozmaitym ustroju fizycznym (a nawet „niefizycznym”) mogą bowiem mieć jednakową strukturę czynnościową” – H. Putnam, *Wiele twarzy realizmu*, w: H. Putnam, *op. cit.*, 1998, s. 338-339.

[210] „Hipotezy stanów funkcjonalnych *nie są* niekompatybilne z dualizmem! (…) System składający się z ciała i «duszy» (…) może równie dobrze być Automatem Probablistycznym” – H. Putnam, *The nature of mental states*, w: H. Putnam, *op. cit.*, 1975, s. 436.

Funkcjonalnych)[211]. Własności poszczególnych stanów umysłowych uznaje się przy tym za cechy drugiego rzędu, które można specyfikować w terminach ról funkcjonalnych odpowiednich cech pierwszego rzędu. Takie ujęcie, proponowane między innymi przez Putnama i Fodora, pozwala przyjąć teoretyczną możliwość dowolnego urzeczywistnienia stanów drugiego rzędu, czyli stanów umysłowych.

Przedstawiciele słabej, fizykalnie redukcjonistycznej wersji funkcjonalizmu uważają, że własności mentalne mogą być specyfikowane wyłącznie za pomocą ról przyczynowych „nośnika" fizycznego. Dzięki tak zwanej *specyfikacji funkcjonalnej* nie redukuje się jednak poszczególnych stanów umysłowych do stanów pierwszego rzędu (np. fizykalnych), natomiast redukuje się je do stanów funkcjonalnych i tym samym istotą stanu umysłowego okazuje się być funkcja stanu fizycznego, a umysł staje się funkcją mózgu. Redukcjonizm polega więc tutaj na stanowczym określeniu materii realizatora własności mentalnych, a właściwie zredukowaniu ich do własności fizycznych, gdzie „poziom funkcjonalny jest tylko pewnym poziomem pojęciowym"[212].

Zwolennicy funkcjonalizmu mocnego „odrywają się" od sfery fizycznej twierdząc, że własności mentalne (np. chęć ucieczki) mogą być specyfikowane za pomocą ról przyczynowych każdego „nośnika" (np. bólu), jako realne własności wyższego rzędu. Teoria superweniencji może stanowić uzupełnienie także tej odmiany funkcjonalizmu, jednak jeśli słaby funkcjonalizm „pojęciowo" wyróżnia tylko dwie warstwy własności (fizyczną i nadbudowaną na niej własność mentalną), to mocny funkcjonalizm ma tendencję do mnożenia tych warstw wraz z określaniem coraz wyższych rzędów własności. Efektem takiego działania jest ulokowanie stanów mentalnych na behawioralnych wyjściach systemu i szukanie odpowiedzi na pytania w rodzaju „czy byka rozjusza *czerwień*, czy *drażliwość* płachty?".

Brak wyjaśnienia, dlaczego stany umysłowe pozostają akurat w takich, a nie innych związkach przyczynowych, pozwala traktować umysł jako swego rodzaju czarną skrzynkę „wypełnioną" relacjami kazualnymi. Dlatego też mocna odmiana funkcjonalizmu bywa nazywana „funkcjonalizmem czarnej skrzynki"[213].

Teoria identyczności egzemplarzy cechuje fizykalizm szczegółowy, zwany też symbolicznym, charakterystyczny dla behawioryzmu logicznego i teorii superweniencji. Teza ta jest również widoczna w słabej odmianie funkcjonalizmu, który ogranicza się tylko do odnalezienia cech fizycznych, posiadających tę samą rolę funkcjonalno-przyczynową, co cechy analizowanego stanu umysłowego. Identyfikacji stanów umysłowych dokonuje się tu poprzez specyfikację w terminach ról funkcjonalnych odpowiednich cech fizycznych.

Specyfikacja funkcjonalna, jako domena funkcjonalizmu analitycznego, jest również wykorzystywana przez psychofunkcjonalistów, którzy ukazują za jej pomocą bardziej

[211] Jest to spadek po behawioryzmie psychologicznym. Teza o identyczności rodzaju (*type-type*) była w nim stosowana w następującym sformułowaniu: „dwa organizmy znajdują się w stanach psychicznych identycznego typu wtedy i tylko wtedy, gdy ich zachowania lub dyspozycje do zachowań są zachowaniami (dyspozycjami) identycznego typu". Behawioryzm był teorią fizykalistyczną, a więc odnosi się do niego również twierdzenie, że „organizmy znajdują się w stanach psychicznych identycznego typu wtedy i tylko wtedy, gdy stany fizyczne, w których się znajdują, są stanami identycznego typu" – cytaty za: J. Fodor, *Czym nie są stany psychiczne?*, w: *op. cit.*, B. Chwedeńczuk (red.), s. 59.
[212] J. Bremer, *op. cit.*, s.114.
[213] J. R. Searle, *op. cit.*, 1999, s. 68.

bezpośrednie relacje między bodźcami, stanami umysłowymi a zachowaniem[214]. Według tej teorii, stan umysłu okazuje się być stanem fizycznym, analitycznie identyfikowalnym dzięki swej roli funkcjonalnej. Rola przyczynowo-funkcjonalna stanów umysłowych staje się wyłącznie czynnikiem pojęciowym i nie jest uznawana za istotę tych stanów. Wynika z tego, że słaby funkcjonalizm, na mocy definicji analitycznych, przyjmuje redukcję własności stanów umysłowych, uznając za ich istotę coś pojęciowego. Stąd nazwa *funkcjonalizmu pojęciowego*.

Funkcjonaliści pojęciowi[215] skłonni są przyjąć, że słowo „ból" jest nieścisłym desygnatorem, konceptualnie równoważnym opisowi w formie „stan z taką a taką rolą przyczynową"[216]. Uznaje się przy tym, że w przypadku człowieka, opisowi temu odpowiada określony stan mózgu, wywołujący stosowne zachowanie. Odpowiednikiem tego opisu *logicznie* mógłby być dowolny stan fizyczny lub nawet nie-fizyczny. Charakterystyczne dla każdej odmiany funkcjonalizmu opisy, które bazują wyłącznie na odniesieniu do wzajemnych relacji przyczynowych stanów z bodźcami i zachowaniami, określane są ukutym przez Johna Smarta mianem „materialnie neutralnych" (*topic-neutral*), ponieważ nie narzucają żadnych logicznych ograniczeń materialnego tworzywa rzeczy, którą opisują.

Można dostrzec pewne podobieństwo w rozróżnieniu mocnego i słabego funkcjonalizmu oraz mocnej Sztucznej Inteligencji (dążącej do stworzenia myślących maszyn) i słabej Sztucznej Inteligencji (nakierowanej na modelowanie ludzkiego mózgu).

Słaba odmiana funkcjonalizmu i Sztucznej Inteligencji spotykają się między innymi na gruncie biocybernetyki i neurofizjologii, czyli nauk mających filozoficzne wsparcie w materialistycznej teorii umysłu. Charakterystyczna dla tej koncepcji krytyka behawioryzmu często opiera się na analogii komputerowej. Michael A. Arbib (biocybernetyk) stwierdza na przykład, że „bez względu na to, w jakiej mierze udał się dobór danych wejściowych i wyjściowych, które uwzględniamy w opisie układu, nie możemy się spodziewać, aby utworzyły one opis kompletny"[217]. Tak, jak bez znajomości programu nie dowiemy się, w co komputer przetworzy dane wejściowe, tak też bez żadnych informacji o stanie czyjejś wiedzy nie jesteśmy zdolni przewidzieć, jak ten ktoś odpowie na określone pytania. Arbib podkreśla, że dla uzyskania pełnego opisu takich układów, jak prosty automat, komputer czy ludzki organizm, niezbędne jest uwzględnienie stanów wewnętrznych tych układów. Opis stanów wewnętrznych powinien się przy tym sprowadzać do określenia, jaki wpływ mają bodźce zewnętrzne (sensoryczne wejścia) na zmianę tych stanów i jakim to skutkuje zachowaniem (behawioralnymi wyjściami). W tym ujęciu, stany mózgu utożsamiane przez Arbiba ze stanami umysłu, są opisywane jako stany funkcjonalne. Podobnie należy traktować stany zewnętrzne, czyli stany otoczenia, którego „wyjścia" są „wejściami", a „wejścia" – „wyjściami" dla danego układu.

Według Arbiba, każdy organizm jest funkcjonalnie związany ze swoim środowiskiem, reagując na „zmienne parametry otoczenia" („zbiór sygnałów wejściowych") i oddziałując na otoczenie „zmiennymi parametrami układu" („zbiorem sygnałów wyjściowych"). „Parametry wewnętrzne układu" (zbiory stanów układu) określają zależności między sygnałami

[214] Psychofunkcjonalizm czerpie w dużej mierze z osiągnięć psychologii kognitywnej. Jak już zauważyłem, przedstawicielem tej odmiany funkcjonalizmu jest m.in. Fodor.
[215] Przedstawicielami funkcjonalizmu pojęciowego są fizykaliści (m.in. David Armstrong i Clarence Lewis).
[216] Dla przykładu: kolor „niebieski" jako *kolor nieba* jest nie-ścisły, jeśli dotyczy dowolnej barwy nieba; kolor „niebieski" jest ścisły, gdy nie określa dowolnej barwy nieba, a jedynie *błękit*.
[217] M. A. Arbib, *Mózg i jego modele*, tłum. S. Bogusławski, PWN, Warszawa 1977, s. 108.

wejściowymi, a wyjściowymi i jako „zmienne parametry układu" mogą, choć nie muszą, oddziaływać na otoczenie.

Arbib uważa, że funkcjonalny opis stanu wewnętrznego stanowi najlepszy opis relacji przyczynowych, zachodzących w przeszłości i przyszłości pomiędzy sygnałami wejściowymi a wyjściowymi. Relacje przyczynowe pomiędzy sygnałami wejściowymi a stanami układu są u Arbiba określane przez tak zwaną „funkcję stanów przejścia", natomiast relacje przyczynowe pomiędzy sygnałami wejściowymi i stanami układu a sygnałami wyjściowymi – przez „funkcję wyjścia".

Jak stwierdza Arbib, pomyślne współoddziaływanie danego układu z otoczeniem jest możliwe tylko wtedy, gdy układ ten posiada szeroki „fundament danych" lub „wewnętrzny model świata", który może być udoskonalany w wyniku funkcjonowania tego układu w środowisku. Arbib głosi tezę, że wynik kontaktów z otoczeniem jest zależny od określonej genetycznie struktury mózgu, stanowiącej podłoże dla mechanizmów pamięci, których działanie przejawia się określonym zachowaniem. Noworodka można więc porównać do komputera z wbudowanymi programami standardowymi, które umożliwiają dopiero programowanie w językach wyższego rzędu. Człowiek, odpowiednio „programowany" swym doświadczeniem, staje się inteligentny[218]. Arbib pojmuje inteligencję jako splot właściwości zachowania, zwanego powszechnie zachowaniem inteligentnym. Sztuczną inteligencję definiuje natomiast jako „takie zaprogramowanie komputera, aby zachowywał się on w sposób, który można nazwać inteligentnym"[219].

Można dopatrywać się tu sugestii, że również człowiek posiada program, który powoduje zachowania inteligentne, i że programem tym jest umysł. Arbib używa cybernetycznego modelu mózgu dla ukazania problemów, które zwykle wchodzą w zakres teorii umysłu (pisze między innymi o postrzeganiu, pamięci, myśleniu i inteligencji). Jest to jednak punkt widzenia naukowca, biocybernetyka i neurofizjologa, który zajmuje się tylko „mechanistycznym" aspektem tych problemów, nie kwestionując „humanistycznej" koncepcji człowieka obdarzonego „rzeczą" zwaną *umysłem*. Gdyby Arbib był filozofem, nie mógłby sobie na taki luksus pozwolić i musiałby albo bardziej dosadnie zakwestionować istnienie umysłu, albo sformułować konkretną propozycję jakiejś teorii umysłu zgodnej z materializmem.

Teorię taką przedstawia David M. Armstrong. Według tego australijskiego filozofa, zadowalającą teorią umysłu może być tylko teoria uzasadniona naukowo, która wyjaśnia jedność ciała i umysłu, zakładając ich interakcję przyczynową. Teoria ta powinna zawierać zasadę numerycznego rozróżniania umysłów i dopuszczać logiczną możliwość ich bezcielesnej egzystencji, zaprzeczając jednak możliwości niezależnego bytu zdarzeń umysłowych. Według Armstronga, materialistyczna teoria umysłu spełnia te wymagania[220].

Armstrong argumentuje, że behawioryzm nie może sprostać wymienionym wymogom, a powodem tego jest głoszone przez zwolenników tej teorii założenie, że wewnętrzne procesy umysłowe po prostu nie mają miejsca. Jeśli przyjmie się tezę o braku wewnętrznych zdarzeń umysłowych między bodźcem a reakcją, to nie może być mowy o interakcji umysłu i ciała.

---

[218] Ważne jest tutaj to, że Arbib nie stara się zredukować człowieka do poziomu działającego komputera, czy robota, lecz używa metafory, która według tego naukowca najlepiej pozwala zrozumieć cechy właściwe tylko człowiekowi. Takie wykorzystanie analogii komputerowej jest charakterystyczne dla słabego funkcjonalizmu.
[219] *Ibidem*, s. 161.
[220] Na podstawie: D. Armstrong, *op. cit.*, s. 54.

Dla behawiorystów umysł nie może oddziaływać na ciało, ponieważ nie jest rzeczą materialną i z tego też powodu za absurd uchodzić może bezcielesna egzystencja umysłu. Armstrong zauważa, że behawioryzm popada w błędne koło, określając umysł w terminach umysłowych, odnoszących się do obserwacji zachowania. Ponadto, behawioryzm nie wyjaśnia wielu ważnych faktów, sprowadzając je do ukrytego zachowania wewnętrznego i, w konsekwencji, do wysyłania przez układ nerwowy impulsów elektrycznych, czego nie można chyba uznać za obserwowalne zachowanie. Armstrong twierdzi, że „zachowanie i dyspozycje do zachowania wchodzą w jakiś sposób w zakres pojęcia umysłu”[221], ale nie wolno identyfikować umysłu z zachowaniem.

Emergentyści twierdzą, że stanom umysłu odpowiadają jedno-jednoznacznie określone stany mózgu, a wszystkie procesy umysłowe „wyłaniają się” z procesów fizycznych, gdy te utworzą odpowiednio złożoną strukturę. Zgodnie z tą koncepcją, nie można teleologicznie wytłumaczyć istnienia jakiegokolwiek procesu umysłowego, który byłby przeciwstawny procesom mózgowym i dlatego od nich niezależny. Konsekwencja ta stanowi, według Armstronga, zarzut wobec emergentyzmu.

Teoria stanu centralnego przyjmuje istnienie stanów umysłowych i identyfikuje je ze stanami mózgu, zaprzeczając tym samym możliwości bezcielesnej egzystencji umysłu. Nie uwzględnianie logicznej możliwości istnienia umysłu poza mózgiem, na przykład, jako układów „elektrycznych wyładowań w przestrzeni”[222], Armstrong uznaje za wadę tej teorii. Twierdzi on, że umysł nie jest „fizyczno-chemicznym wytworem centralnego układu nerwowego”[223], a może być uznawany za mózg tylko w takim sensie, w jakim gen uznaje się za cząsteczkę DNA (umysł jest opisem mózgu tak samo, jak gen stanowi opis cząsteczki DNA na poziomie funkcjonalnym).

Według Armstronga, niektóre stany umysłowe mogą być przyczyną zachowania. Jest to dla tego filozofa oczywiste z uwagi na fakt, że „przyczynowość w królestwie umysłu nie różni się niczym od przyczynowości w królestwie fizyki”[224].

W celu ukazania procesów zachodzących w umyśle, Armstrong odnosi się do procesów przyczynowych zachodzących w mózgu, opierając się przy tym na analogii komputerowej. Wcześniej jednak formułuje *przyczynową teorię inferencji*, uznając twierdzenie „*A* wnioskuje *p* z *q*” za równoznaczne twierdzeniu „sąd *A*, że *q*, *skłania A* do przyjęcia sądu, że *p*”, co właściwie oznacza, iż „sąd *A*, że *q* jest *przyczyną* przyjęcia przez *A* sądu *p*” [225].

Według Armstronga, proces będący odpowiednikiem wnioskowania zachodzi zarówno w żywym mózgu, jak i w komputerze. Polega on na przejściu do nowego stanu, będącego przyczynowym skutkiem stanu poprzedniego. Armstrong sugeruje, że sama inferencja nie jest już procesem, lecz *zdarzeniem* umysłowym, polegającym na nabyciu nowego przekonania, które jest przyczynowym skutkiem przekonań nabytych wcześniej, będących składnikami wiedzy inferencyjnej lub nieinferencyjnej. Przekonania można bowiem nabywać nie tylko w drodze wnioskowania, ale także jako postrzeżenia. Postrzeżenie koloru jakiejś rzeczy jest dla Armstronga nabyciem przekonania, że ta rzecz ma określony kolor, czyli nabyciem zdolności

[221] *Ibidem*, s. 108.
[222] *Ibidem*, s. 107.
[223] *Ibidem*, s. 126.
[224] *Ibidem*, s. 118.
[225] Armstrong wzoruje się przy tym na teorii inferencji sformułowanej przez Hume’a.

do określonego zachowania w stosunku do rzeczy danego koloru. Przekonanie postrzeżeniowe, jako zdolność do określonego zachowania, może pojawić się tylko dzięki istnieniu rzeczy mających na przykład odpowiedni kolor. Armstrong zaznacza, że rzeczy nie wyglądają na kolorowe, lecz *są* kolorowe, a ich własności wtórne nie są niczym więcej, niż własnościami fizycznymi tych rzeczy. Dotyczy to również własności umysłowych.

Filozof ten uważa, że przekonania postrzeżeniowe są stanami umysłu tworzącymi strukturę o zmiennej ciągłości egzystencji. Struktura ta układa się w mniej lub bardziej adekwatną „mapę" rzeczywistości (nie tylko otoczenia, ale także ciała i umysłu). Jest to „mapa", która „z samej swej natury wskazuje na fizyczny stan rzeczy, jaki odtwarza"[226]. Równie złożona jest struktura zdań, za pomocą których przekonania mogą być wyrażane. Zdania te są w różnym stopniu skorelowane z „mapą" świata, którą można sobie wyobrazić jako sieć pokrywającą jakąś fizyczną powierzchnię.

Zgodnie z myślą Armstronga, strukturę lingwistyczną można uznać za sieć w jakimś stopniu izomorficzną do sieci wszystkich przekonań, przekonaniami „niepostrzeżeniowymi" (np. inferencyjnymi). Struktura tych ostatnich składa się bowiem z takich samych elementów, jak wyżej opisana „mapa" rzeczywistości. Armstrong uzasadnia to twierdzeniem, że przekonania „niepostrzeżeniowe" dają się ostatecznie sprowadzić do pojęć postrzeżeniowych, których treść konstytuowana jest przez ich związek z rzeczywistością.

Według Armstronga, nabywanie przekonań ściśle wiąże się z myśleniem. Aby to wyjaśnić, filozof ten ponownie posługuje się analogią komputerową. Stwierdza mianowicie, że informacje przekazywane do komputera mogą być przyczynowo czynne lub nieczynne. Informacja jest przyczynowo nieczynna wtedy, gdy zakodowana jest w pamięci komputera i nie odgrywa żadnej roli w jego funkcjonowaniu. Armstrong porównuje ten stan do stanu posiadania przez człowieka jakiejś wiedzy, która nie skutkuje zmianą stanu umysłowego. Jeśli wiedza ta staje się przyczynowo czynna i powoduje zmiany stanu umysłu, to mamy do czynienia z myśleniem. Myślenie może być przy tym równoznaczne z nabywaniem nowych przekonań, choć można je nabyć również prawie bezmyślnie. Prawie, gdyż o obecności myśli decyduje nawet najprostsze oddziaływanie na umysł przekonań postrzeżeniowych. Myślenie może przejawiać się w określonym zachowaniu, jednak nie należy go wiązać wyłącznie z potencjalnym zachowaniem słownym. Myśli bowiem wszystko to, co posiada przekonania wpływające na zachowanie tego czegoś.

Armstrong stwierdza, że „w umyśle mamy do czynienia z celowym łańcuchem myśli, w którym każda z następujących po sobie powstaje w rezultacie manipulacji myślą poprzednią, zgodnie z jakąś ustaloną regułą", przy czym „w wielu przypadkach rozważanie takie zawiera inferencję: nabywanie nowych przekonań jako przyczynowy rezultat przekonań poprzednio posiadanych"[227]. Filozof ten uważa, że wyjaśnienie zdarzeń, procesów i stanów umysłowych musi odwoływać się do fizycznego oddziaływania na ciało i jego zachowania. Stany umysłowe, będące według Armstronga fizyczno-chemicznymi stanami mózgu, powinny być wyjaśniane wyłącznie w terminach skutków oddziaływania otoczenia i przyczyn określonego zachowania.

Powyższe twierdzenia można sparafrazować zdaniem, że stany umysłowe są stanami fizycznymi, stanowiącymi przyczynowe pośredniki pomiędzy sensorycznymi wejściami i

[226] *Ibidem*, s. 474.
[227] *Ibidem*, s. 487.

behawioralnymi wyjściami. W ten właśnie sposób, identyfikując stany umysłowe z fizycznymi i określając je za pomocą ich ról przyczynowych, Armstrong godzi teorię psychofizycznej identyczności z funkcjonalizmem osłabionym tezami redukcjonistycznymi.

2.2.4. Kłopoty z funkcjonalizmem

Funkcjonalizm sprowadza stany umysłowe danego systemu do relacji przyczynowych, zachodzących między tymi stanami oraz między nimi a wejściami i wyjściami tego systemu. Twierdzenie o wielorakiej realizacji stanów umysłowych mówi, że wyżej wspomniane relacje przyczynowe mogą zostać odtworzone w każdym systemie o odpowiednich własnościach przyczynowych, bez względu na to, czy będzie to struktura ludzkich neuronów, czy komputerowych układów scalonych. Twierdzenie to otworzyło drogę funkcjonalizmowi komputerowemu, który jako ideologia Sztucznej Inteligencji, głosi między innymi tezę, że realizacja odpowiedniego programu komputerowego, z właściwymi wejściami i wyjściami, jest wystarczająca dla pojawienia się stanów umysłowych. Aby uczynić tę ideologię bardziej wiarygodną, poszukuje się naukowego uzasadnienia dla zaistnienia stanów umysłowych i tworzy nowe nurty w psychologii (psychofunkcjonalizm).

Jak zauważa John R. Searle, nauka jest obiektywna, ponieważ „zajmuje się rzeczywistością, która jest obiektywna”[228]. Według funkcjonalizmu naukowego, stany umysłowe należałoby badać obiektywnie, z punktu widzenia trzeciej osoby, uznając je za przyczynowe pośredniki między obserwowalnymi bodźcami (wejściami) a obserwowalnym zachowaniem (wyjściami) myślącego systemu.

Podobne założenia tkwią u podstaw behawioryzmu, który w odróżnieniu od funkcjonalizmu, nie widział niczego między bodźcami a zachowaniem, co mogłoby odgrywać jakąkolwiek rolę przyczynową. Dyspozycje do zachowania są w behawioryzmie jedynie wrodzonym elementem procesu reagowania na określone bodźce (jedyną przyczyną jest bodziec, którego jedynym skutkiem może być odpowiednia reakcja).

Nie bez powodu więc funkcjonalizm bywa często nazywany „ulepszoną wersją”[229], czy „reinkarnacją behawioryzmu”[230], co oznacza, że można wobec obu tych koncepcji przedstawić podobne zarzuty, jeśli podważy się „dogmaty tradycji”, na której wyrosły, a jest to tradycja materialistyczna[231]. Obiektywna „rzeczywistość naukowa” jest bowiem rzeczywistością fizyczną, opisywaną w kategoriach przyczyny i skutku. Jeśli więc ontologia umysłu ma być uzasadniona naukowo, to musi to być ontologia kauzalna (odnosząca się do „rzeczywistości naukowej”). Epistemologia umysłu powinna natomiast być epistemologią behawioralną (spełniającą wymóg naukowej obiektywności).

Głoszona przez zwolenników mocnego funkcjonalizmu zupełna dowolność realizacji własności umysłowych pozwala trzymać się z dala od głoszenia twierdzeń ontologicznych, ale nie uwalnia od tradycji materialistycznej. Tradycja ta odziedziczyła po kartezjanizmie aparaturę pojęciową, zgodnie z którą „umysłowe” implikuje „nie-fizyczne”, a „fizyczne”

[228] J. R. Searle, *op. cit.*, 1999, s. 26.
[229] Paul Snowdon, *Funkcjonalizm*, w: *op.cit.*, T. Honderich (red.), t. I, s. 287.
[230] N. Block, *Troubles with functionalism*, za: U. Żegleń, *op. cit.*, s. 62.
[231] W sprawie behawioryzmu i funkcjonalizmu Searle wypowiada się następująco: „akceptacja poglądu, że między umysłem a zachowaniem zachodzi pewien istotny związek, jest wspólną cechą każdego z tych stanowisk” - J. R. Searle, *op. cit.*, 1999, s. 27.

implikuje „nie-umysłowe”. Aby uniknąć sporów wynikłych z kartezjańskiego dualizmu, materializm uznał to, co „nie-fizyczne” za „nie-obiektywne”, czyli „nie-naukowe” i dlatego „nie-rzeczywiste”, eliminując ze swego słownika pojęcie „umysłowości” jako kategorię psychologii potocznej (mowa tu o materializmie eliminacyjnym). Postawa taka jest jaskrawo widoczna w behawioryzmie, ale jej echa mają wpływ na funkcjonalizm, szczególnie na jego słabą odmianę, w której stany mentalne są tylko epistemologicznie identyfikowane dzięki swej funkcjonalnej roli, identycznej z funkcjonalną rolą odpowiedniego stanu fizycznego.

Mocny funkcjonalizm nie określa otwarcie statusu ontologicznego własności pierwszego rzędu, ale utrzymuje w mocy tradycyjne, kauzalne ujęcie umysłu, wsparte wymogiem obiektywności badań (wymóg ten spełniają między innymi zasady testu Turinga).

Searle określa własności umysłowe jako emergentne własności wyższego rzędu, których bazą są własności fizyczne, a właściwie biologiczne (filozof ten określa swe stanowisko mianem *naturalizmu biologicznego*). Według Searle’a, stan umysłowy jest stanem fizycznym takiego tworu biologicznego, jak mózg, i jest nim w tym samym sensie, w jakim ciekły stan skupienia jest stanem takiego systemu molekuł, który tworzy wodę. Uznanie własności umysłowych za „wyłaniające” się z własności biologicznych zakłada warstwową budowę rzeczywistości, zaprzeczając adekwatności tradycyjnego aparatu pojęciowego („umysłowe” nie implikuje tu „nie-fizycznego” i „nie-rzeczywistego” – własności umysłowe są dla Searle’a równie rzeczywiste, jak płynność wody).

Searle uważa, że perspektywa trzeciej osoby (wymóg obiektywności) dotyczy wyłącznie kwestii przyczynowości (zachowania, funkcjonowania)[232] oraz epistemologii (w sensie poznania „obiektywnego”)[233] i nie powinna być rozszerzana do sfery ontologicznej, gdzie umysł może być adekwatnie ujęty jedynie z perspektywy pierwszej osoby.

Odpowiedź na pytanie: *co to jest umysł?* kryje się, według Searle’a, w „ontologii z punktu widzenia pierwszej osoby”[234]. Wyżej wspomniane rozszerzenie jest przyczyną utożsamiania inteligencji z inteligentnym zachowaniem i definiowania umysłu tylko w sposób „obiektywny”, na podstawie obserwowalnego funkcjonowania jakiegoś systemu[235]. Jak stwierdza Searle, „nie cała rzeczywistość jest obiektywna; część rzeczywistości jest subiektywna”[236], a subiektywność jest nieodłącznym atrybutem umysłu.

Searle uznaje, że umysł składa się z *qualiów*, czyli jakościowych stanów mentalnych, które są zjawiskami rzeczywistymi, ale subiektywnymi i jako takie nie mogą być dostępne „trzeciej

[232] W dziedzinie przyczynowości mieści się na przykład funkcjonalna definicja serca, jako organu pompującego krew.
[233] Może tu chodzić na przykład o poznanie organu pompującego krew, czyli serca, które można badać na przykład za pomocą elektrokardiografu.
[234] Searle sugeruje, że właściwą odpowiedzią na pytanie *Co to jest serce?* nie jest stwierdzenie, że „serce to organ pompujący krew, który można badać na przykład za pomocą elektrokardiografu” – jest to definicja, którą można by nazwać „funkcjonalno-epistemiczną”, tymczasem, według Searle’a, właściwą odpowiedzią na zadane wcześniej pytanie może być tylko definicja „ontologiczna”, zgodnie z którą „serce to duży narząd zbudowany z tkanki mięśniowej, mieszczący się w klatce piersiowej”. Searle twierdzi, że *umysł* również należy zdefiniować „ontologicznie”.
[235] „Uporczywa tendencja do formułowania pytania: «W jakich warunkach skłonni jesteśmy przypisywać komuś stany mentalne?» doprowadziła nas do behawioryzmu, funkcjonalizmu, mocnej wersji SI, materializmu eliminującego (…) i niewątpliwie do różnych innych nieporozumień” – *Ibidem*, s. 35.
[236] *Ibidem*, s. 38.

osobie". Stany umysłowe są zawsze „czyjeś", co wskazuje, że „punkt widzenia pierwszej osoby jest pierwotny"[237].

Ujawnia się tu jeden z podstawowych problemów funkcjonalizmu (a zarazem Sztucznej Inteligencji) oparty na argumencie nieobecności qualiów. Argument ten można streścić w następujący sposób: funkcjonalizm jest fałszywy, gdyż możliwe jest wyobrażenie sobie czegoś, co jest funkcjonalnie takie, jak my, a mimo tego nie posiada qualiów[238].

Podsumowaniem dotychczasowych rozważań może być stwierdzenie, że tradycyjne, obiektywistyczne ujęcie tego, co umysłowe (ujęcie z punktu widzenia trzeciej osoby), jest błędne, a jeśli nie błędne, to co najmniej niewystarczające. Nie można formułować twierdzeń o istnieniu stanów umysłowych jakiegoś systemu wyłącznie na podstawie obserwacji zachowania tego systemu (na przykład zachowania językowego, jak w przypadku testu Turinga).

Jak twierdzi Searle, przyczynowość i epistemologia *pomagają* tylko wykryć pewną „bazę" ontologiczną, określająca naturę stanów umysłowych. Zgodnie z funkcjonalizmem, „natura" stanu umysłowego jest taka, jak natura automatu: ukonstytuowana przez relacje do innych stanów oraz wejść i wyjść. „Naturę" pojętą w ten sposób można w pełni scharakteryzować w języku logiczno-matematycznym i poprzez określenia właściwe dla sensorycznych wejść i behawioralnych wyjść. Nie mamy tu jednak do czynienia z wykryciem prawdziwej „bazy" ontologicznej stanu umysłowego, lecz z przeformułowaniem ontologii w kategoriach przyczynowości i epistemicznych podstaw perspektywy trzeciej osoby. Inaczej mówiąc: stwierdzając, że istotą stanu umysłowego jest jego funkcja (rola przyczynowa), funkcjonaliści nie opuszczają perspektywy trzeciej osoby i formułują tezy „ontologiczne", nie wychodząc poza dziedzinę nauki i takiej epistemologii, w której podmiot jest obserwatorem poznającym tylko rzeczywistość obiektywną.

Searle głosi pogląd, że można w ten sposób poznać tylko *przejawy* umysłowości, należącej do rzeczywistości subiektywnej, z którą ani materializm, ani behawioryzm i funkcjonalizm nie mogą sobie poradzić. Rzeczywistość nie-naukowa nie jest rzeczywistością behawiorystów i funkcjonalistów, a sama nauka nie da pełnego wyjaśnienia, czym jest umysł. Dlatego nauce o umyśle (a właściwie o tym, jak się umysł fizycznie objawia) musi towarzyszyć ujęcie metafizyczne, czyli filozofia umysłu.

Problem ukazany przez Searle'a może też zostać ujęty w inny sposób. Można mianowicie przyjąć, że obiektywnie istniejąca rzeczywistość to *zewnętrzny* kosmos energii, materii i ciał. Opisem tej rzeczywistości zajmuje się nauka i ontologia, które cechuje przyjęcie punktu widzenia trzeciej osoby i metodologiczna eliminacja podmiotu. Searle uznał, że epistemologia i ontologia mogą ujmować rzeczywistość „pierwszoosobowo" i „trzecioosobowo", przy czym umysł należy do rzeczywistości w ujęciu pierwszym, objawiając się jedynie w rzeczywistości ujętej z punktu widzenia trzeciej osoby.

W obecnie prezentowanym ujęciu[239], ontologiczna konceptualizacja rzeczywistości jest *bezpodmiotowa*. Podmiot jest w niej obecny tylko funkcjonalnie, jako czynnik realizacji poznania. W ontologii i nauce liczy się bowiem tylko relacja poznania do obiektywnej

[237] *Ibidem*, s. 40.
[238] Argument nieobecności *qualiów* został przedstawiony przez Neda Blocka. Szczegółowe omówienie tego argumentu nie jest tutaj konieczne.
[239] Jest to ujęcie zaczerpnięte z prac Andrzeja Chmieleckiego.

rzeczywistości, a anonimowy podmiot pełni funkcję „nośnika" treści poznania. Można tu mówić o poznaniu intersubiektywnym, bezpodmiotowym, opartym na kumulacji treści wyrażanych przez różne dowolne osoby.

Gdy przyjmie się graniczne pojęcie podmiotu, jako podmiotu aktów duchowych, stojącego w opozycji wobec czegokolwiek, co może być uznane za przedmiot, wtedy nie może być mowy o funkcjonalnym uprzedmiotowieniu podmiotu poznania[240]. Można by rzec, że podmiot *Poznania* staje się *Podmiotem* poznania, a samo poznanie odchodzi na dalszy plan.

Spojrzenie na podmiot z punktu widzenia pierwszej osoby („uczestnika", a nie „obserwatora"), oznacza wejście w obręb rozważań metafizycznych. Metafizyka jest teorią rzeczywistości irrealnej, czyli kosmosu *wewnętrznego*, ukonstytuowanego przez podmiot duchowy. Teoria ta nie dotyczy jednak jakiegoś określonego kosmosu wewnętrznego, czyli światopoglądu konkretnego podmiotu aktów duchowych, lecz kosmosu wewnętrznego jako takiego. Ta właśnie perspektywa stanowi dopełnienie konceptualizacji naukowych i ontologicznych, ponieważ ontologiczny model rzeczywistości realnej powinien zawierać się w metafizycznym modelu rzeczywistości irrealnej. Trzeba bowiem zdawać sobie sprawę, że konceptualizacja obu tych modeli należy do rzeczywistości wirtualnej, czyli kosmosu wewnętrznego, który jest prawdziwy na tyle, na ile jest izomorficzny z kosmosem zewnętrznym.

Zarówno filozof zajmujący się metafizyką, jak i naukowiec z filozofem zajmującym się ontologią, badają rzeczywistość. Wszyscy oni poszukują rzeczywistego umocowania bytowego fenomenów. Drogi ustalania determinant danych zmysłowych są jednak różne. Naukowiec i ontolog szukają ich umocowania bytowego w determinantach przedmiotowych (transcendentnych), zastępując to, co subiektywne tym, co obiektywne[241], natomiast metafizyk skupia się na determinantach podmiotowych (transcendentalnych).

Metafizyka nie miałaby aspektu poznawczego, gdyby nie liczyła się z naukową i ontologiczną konceptualizacją obiektywnie istniejącej rzeczywistości. Ale też sama nauka (dążąca do ustalenia, *jaka* jest rzeczywistość), czy sama ontologia (mająca określić sposób bycia rzeczywistości – *jak* ona jest), nie są w stanie dostarczyć pełnego ujęcia rzeczywistości. Tymczasem to właśnie na gruncie nauki i ontologii wyrosła teoria funkcjonalistyczna, a wraz z nią ideologia mocnej odmiany Sztucznej Inteligencji, dziedzicząca problemy, z którymi zmagali się behawioryści i fizykaliści, a po nich funkcjonaliści.

Jeśli samo źródło tych problemów zostało już ukazane, to należy przejść do przedstawienia tego, co z tego źródła „wypłynęło".

Funkcjonalizm różni się od behawioryzmu głównie tym, że formułuje twierdzenia o wzajemnych relacjach przyczynowych stanów umysłowych, jednak tak samo, jak behawioryzm, jest obciążony tym, co Ned Block nazwał *liberalizmem*. Wynikająca z założeń funkcjonalizmu teza o wielorakiej realizacji stanów umysłowych pozwala przypisywać te stany rzeczom, które ich nie mają (głowom wypełnionym homunkulusami, całym narodom, czy komputerom). Ten liberalizm otworzył drzwi do „chińskiego pokoju" – słynnego argumentu Johna Searle'a, podważającego ideologię Sztucznej Inteligencji[242].

[240] Koncepcja podmiotu aktów jako czystego epistemologicznego podmiotu transcendentalnego pochodzi od Heinricha Rickerta (1863-1936), urodzonego w Gdańsku przedstawiciela niemieckiego neokantyzmu.
[241] Na przykład redukując doznanie bólu do pobudzeń włókien nerwowych C_4.
[242] Argument ten zostanie przedstawiony w następnym rozdziale.

Sposobem na liberalizm może być albo „osłabienie" funkcjonalizmu poprzez zbliżenie go do fizykalizmu, czyli teorii identyczności stanów umysłowych ze stanami mózgu, albo też poprzez uznanie funkcjonalizmu za empiryczną teorię naukową (tak powstał psychofunkcjonalizm). Funkcjonaliści uznają jednak fizykalizm za teorię *szowinistyczną*, która odmawia posiadania stanów umysłowych systemom, o których intuicyjnie można stwierdzić, że je mają. Słaba odmiana funkcjonalizmu nie chroni się przed szowinizmem nawet wtedy, gdy stanowi tylko swego rodzaju logiczny „instrument", który pozwala takim materialistom, jak Armstrong, analizować pojęcia umysłowe poprzez odwołanie się do ról przyczynowych stanów fizycznych. Funkcjonalizm jest „skażony" szowinizmem pośrednio, przez fizykalizm, albo bezpośrednio, jako naukowa teoria psychologiczna.

Jak zauważył Searle, funkcjonalizm opiera się na przyczynowości. Jeśli własności umysłowe uzna się za własności drugiego rzędu jakichś innych własności, spełniających pewne warunki przyczynowe, to własności pierwsze okazują się być epifenomenami własności drugich. Jeżeli własności umysłowe są tylko epifenomenami, to jak mogą oddziaływać *przyczynowo*?

Problem przyczynowania mentalnego w odniesieniu do funkcjonalizmu można ukazać w prosty sposób: własności umysłowe są drugorzędnymi własnościami względem własności ich neuronalnych realizatorów i to one przejmują moce przyczynowe własności umysłowych. Ale to nie koniec. Własności neuronalne są przecież drugorzędnymi własnościami względem jakichś własności niższego rzędu, również przejmujących moce przyczynowe swoich własności nadrzędnych. Trudno stwierdzić, czy ta „redukcja" może mieć koniec, a przecież jego brak zaprzeczyłby istnieniu jakiejkolwiek przyczynowości. Można przypuszczać, że bazowych własności przyczynowych doszukiwano by się obecnie gdzieś na poziomie mikrofizycznym. Mikrofizyka okazałaby się zatem jedyną nauką o mocach przyczynowych. Gdyby tak było, to własności umysłowe nie mogłyby oddziaływać przyczynowo tak samo, jak rzucony kamień nie mógłby być przyczyną rozbicia szyby.

Jak zauważa Jaegwon Kim, problem tkwi w tym, że własności umysłowe i ich własności bazowe są w funkcjonalizmie uznawane za własności tej samej rzeczy, należącej do jednego poziomu w hierarchii mikro-makro[243]. Kim uważa, że poziomów ontologicznych jest wiele i odwołuje się przy tym do teorii superweniencji. Jest on funkcjonalistą pojęciowym, gdyż twierdzi, że funkcjonalne własności wyższego rzędu superweniują na pierwszorzędnych własnościach ich bazowych realizatorów, co może mieć miejsce tylko na jednym, *pojęciowym* poziomie ontologicznym[244].

Funkcjonalizm postuluje istnienie wielu rzędów własności, nie rozróżniając poziomów ontologicznych, na których one występują. Stąd biorą się trudności przyczynowego wykluczania. Jak twierdzi Kim, *„makrowłasności mogą posiadać, i w ogóle posiadają swoje własne moce przyczynowe, które przekraczają moce przyczynowe ich mikroskładników*"[245]. Nie można w związku z tym dokonywać „przyczynowej" redukcji makrowłasności (takich, jak własności umysłowe) do mikrowłasności (takich, jak na przykład własności fizyczne).

[243] Dla reprezentantów słabej odmiany funkcjonalizmu poziomem tym mógłby być tylko poziom fizyczny (lub fizyko-chemiczny), a dla przedstawicieli mocnego funkcjonalizmu poziom ten nie jest najważniejszy i mogłaby tu wchodzić w grę nawet substancja duchowa.

[244] „Dobroć Sokratesa superweniuje na jego uczciwości, hojności, odwadze oraz mądrości. Jednak to ta sama osoba, Sokrates, egzemplifikuje zarówno subwenientne cnoty, jak i superwenientą dobroć". Sokrates jako osoba należy do określonego poziomu ontologicznego, innego, niż Sokrates jako organizm biologiczny i innego, niż Sokrates jako struktura jakichś fizykalnych mikroskładników (cytat z: J. Kim, *op. cit.*, s.96).

[245] *Ibidem*, s. 96.

Mogłoby to sugerować, że Kim dopuszcza istnienie rzeczywistości mentalnej, rozumianej jako rzeczywistość nie-fizyczna. Kim jest jednak fizykalistą, więc rozróżnienie poziomów ontologicznych jest u niego rozróżnieniem czysto pojęciowym. Sugeruje, aby termin „własności drugiego rzędu" zastąpić „opisami", „desygnatorami", czy po prostu „pojęciami drugiego rzędu".

Według tego filozofa, świat jest światem fizycznym, w którym niektórych zjawisk nie da się fizykalnie wyjaśnić. Zjawisk tych nie można wyjaśnić za pomocą pojęć opisujących własności pierwszego rzędu i dlatego „pojęcia drugiego rzędu", grupujące własności rzędu pierwszego poprzez specyfikację funkcjonalną, są praktycznie niezbędne dla celów *komunikacyjnych*, choć same w sobie nie reprezentują żadnych nowych własności.

„Pojęcia drugiego rzędu" nie są adekwatne zawsze, ale tylko w odpowiednim kontekście ich użycia. Przede wszystkim należy szukać odpowiedzi na pytanie, *dlaczego* zachodzi korelacja między zjawiskami wyjaśnianymi fizykalnie, a własnościami umysłowymi, rozumianymi jako „pojęcia drugiego rzędu". W drodze tych poszukiwań, Kim proponuje zastosować redukcję funkcjonalną, która jest możliwa tylko po dokonaniu *funkcjonalizacji* własności umysłowych[246].

Według funkcjonalizmu, własności mentalne mogą posiadać wiele różnych realizatorów. Kim przyjmuje, że mogą to być tylko realizatory fizyczne, zaś wieloraka realizacja dotyczy różnych gatunków i struktur w różnych nomologicznie możliwych światach. Redukcja własności umysłowych do fizycznych jest w takim przypadku relatywna względem gatunków, struktur i światów odniesienia, przy czym własności umysłowe „dziedziczą" moce przyczynowe po swych realizatorach. Jeśli więc jakaś własność umysłowa *U* jest własnością fizyczną F_1 w gatunku *1*, F_2 w gatunku *2* i F_3 w gatunku *3*, to własność ta wielorako realizuje się w różnych własnościach *Fi*.

Konkluzja Kima jest jednoznaczna. Jeśli przyjmie się, że własności umysłowe mogą podlegać funkcjonalizacji, to należy również udzielić odpowiedzi, w jaki sposób poddać funkcjonalizacji własności jakościowe (*qualia*). Jeśli nie znajdzie się na to odpowiedzi, to pozostaje albo przyjęcie dualizmu własności, albo dualizmu substancjalnego. Innym wyjściem jest zostać fizykalistą, ale takim, który nie redukuje własności mentalnych do fizycznych (monizm anomalny). Można jednak pójść dalej i stwierdzić, że umysł redukuje się do tego, co fizyczne (materializm redukcjonistyczny). Sfera mentalna zostaje w ten sposób pozbawiona jakiejkolwiek mocy przyczynowej (epifenomenalizm), co według Kima prowadzi do mentalnego irrealizmu (fizykalizm eliminacyjny). Filozof ten proponuje, aby uznać całą sferę mentalną za realną część świata fizycznego, którą można częściowo (oprócz *qualiów*) wytłumaczyć za pomocą „pojęć drugiego rzędu", w drodze funkcjonalizacji własności, czyli specyfikacji funkcjonalnej.

Podejście Kima jest typowe dla słabego funkcjonalizmu (a konkretnie dla funkcjonalizmu pojęciowego), którego przedstawicielem był również Armstrong. Kim sprzeciwia się mocnemu funkcjonalizmowi, jako odmianie antyredukcjonizmu w sprawie umysłu i ciała. Filozof ten stwierdza, że „redukcja umysłu i ciała *wymaga* funkcjonalistycznej koncepcji własności mentalnych" i trafnie zauważa, że „jeśli jest to pogląd słuszny, to redukcjonizm

[246] Funkcjonalizacja własności umysłowych polega na zdefiniowaniu tych własności w terminach ich przyczynowych relacji do innych własności ze względu na obowiązujące w danym świecie prawa. Relacje te są więc metafizycznie przygodne, gdyż zmieniają się w zależności od światów, w których zachodzą.

umysłu i ciała oraz funkcjonalistyczne podejście do tego, co mentalne, podzielają ten sam metafizyczny los broniąc się i upadając razem”[247].

Rzeczywiście – próby przyczynowego wyjaśniania wszystkich zjawisk (nawet poprzez funkcjonalizację własności niewyjaśnialnych fizycznie) pogrążają zarówno fizykalizm, jak i funkcjonalizm. Fizykalizm opiera się na twierdzeniach naukowych i ontologicznych (jako teoria materialistyczna, redukuje to, co umysłowe do tego, co może wyjaśnić fizyka). Funkcjonalizm z kolei opiera się na przyczynowym wyjaśnieniu samej *istoty* umysłu. Pomija w ten sposób kwestię *natury* umysłu, formułując jakby „niepełne” twierdzenia ontologiczne. Jeśli bowiem uznaje się, że umysł jest bytem, to musi mieć jakąś naturę. Natura umysłu może zostać określona poprzez ukazanie jego miejsca w strukturze rzeczywistości. Można tego dokonać, znajdując czynniki realizacji (czyli określając, w czym umysł jest realizowany i jak powstaje) i czynniki determinacji (ustalając, jakim prawidłowościom umysł podlega i od czego zależą cechy *istotowe* umysłu, czyli to, że umysł jest umysłem). Funkcjonalizm ogranicza się tylko do twierdzenia, że umysł jest umysłem ze względu na określone związki typu funkcjonalnego.

Funkcjonalista odpowiada na pytania o istotę umysłu, pojmując determinizm tak, jak materialiści, jako oddziaływanie przyczynowo-skutkowe. Nie może sobie jednak poradzić z problemami przyczynowania mentalnego, ponieważ uważa, że głosi teorię natury umysłu, mówiąc tak naprawdę tylko o istocie umysłu (i to błędnie!). Stara się określić istotę umysłu (umocowanie bytowe umysłu) w kategoriach zarezerwowanych dla fizycznych czynników realizacji (bytowego fundamentu umysłu), które mogą oddziaływać przyczynowo. Mogłoby się wydawać, że tacy reprezentanci słabego funkcjonalizmu, jak Kim, dostrzegli ten problem, ale jest to złudne. Kim, jako fizykalista, rzeczywiście mówi o naturze umysłu. Odwołuje się w tym do teorii superweniencji i twierdzi, że jej uzasadnieniem jest funkcjonalizm. Mamy tu dwie niepełne teorie ontologiczne, które według Kima wzajemnie się uzupełniają. Jednak tak, jak funkcjonalizm błędnie ujmuje istotę umysłu, tak też teoria superweniencji dotyczy natury umysłu tylko w zakresie fizycznej jego realizacji.

Czy jednak tworzywo umysłu ma znaczenie? Jak zauważa Daniel C. Dennett, umysł może zostać zrealizowany tylko w układzie, który musi być wyposażony w liczne przetworniki i efektory, i w którym informacja musi być przekazywana odpowiednio szybko. To, że „nie moglibyśmy skonstruować *czującego* umysłu z krzemowych układów scalonych, drutu, szkła czy puszek po piwie powiązanych sznurkiem”[248], nie jest według Dennetta powodem do odrzucenia funkcjonalizmu. Aby „nośniki” umysłu mogły spełniać swoje funkcje, muszą być zbudowane z takiej materii, która jest *biohistorycznie* „kompatybilna” z kontrolowanymi przez te „nośniki” ciałami. Jeśli to, że „nośniki” te mogą funkcjonować jako składniki większych układów funkcjonalnych, zależy od tego, z czego te „nośniki” są zbudowane, to błędem by było pomijać ich naturę. Dennett twierdzi jednak, że natura „nośników” umysłu ma takie samo znaczenie, jak wódka czy narkotyk, a przecież „nie ma więcej gniewu czy lęku w adrenalinie niż głupoty w butelce whisky”[249].

Według Dennetta problem tkwi w tym, że funkcjonaliści zbyt upraszczają swoją koncepcję, pozornie tylko „ułatwiając sobie życie”. Nie można bowiem porównywać układu nerwowego do struktury komputera. Dennett zauważa, że do urządzeń wejściowych komputera zalicza się wszystkie przetworniki, które przekładają informacje pochodzące ze środowiska

[247] *Ibidem*, s. 111.
[248] D. C. Dennett, *op. cit.*, s. 92.
[249] *Ibidem*.

(zewnętrznego lub wewnętrznego) na jedno „wspólne medium przetwarzania informacji". Zarysowuje się tu widoczna granica pomiędzy kanałami informacji w komputerze, a jego środowiskiem.

Dennett sądzi, że nie można symulować komputerowo interakcji zachodzących we wszystkich przetwornikach i efektorach układu nerwowego. Odizolowanie kanałów informacyjnych od „wejściowych" zdarzeń zachodzących w układzie nerwowym jest niemożliwe ze względu na to, że układ ten jest po prostu zbyt skomplikowany i składa się ze zbyt wielu efektorów i przetworników, występujących prawie na każdym złączu. Tymczasem funkcjonalistyczno-komputacyjna teoria sztucznej inteligencji sprowadza architekturę umysłu do maszyny Turinga, realizowanej w komputerze cyfrowym, który składa się z procesora centralnego, pamięci, wejść i wyjść. Współczesny naukowiec nie może pozwolić sobie na taką analogię. Dlatego neuronauka przyjmuje obecnie koneksjonistyczną architekturę umysłu, opartą na równoległym i rozproszonym przetwarzaniu informacji.

Jak zauważa Dennett, w układzie nerwowym nie ma procesora centralnego. Informacje ze środowiska, dostarczane za pomocą narządów zmysłowych, są przetwarzane równolegle, w wielu różnych strukturach komórek nerwowych i obszarach mózgu. Gdyby mózg uznać za komputer, to informacje musiałyby być w nim gromadzone w postaci odrębnych plików, umieszczanych w pamięci pod odpowiednim adresem[250]. Zgodnie z odkryciami współczesnej neuronauki, informacje są wprawdzie zakodowane w mózgu i przechowywane w pamięci, ale zdolność zapamiętywania ma naturę czysto biologiczną, opartą na syntezie białka, zapisie genetycznym i elektrycznej stymulacji mózgu. Nie może tu być mowy o plikach, czy adresach.

Układu nerwowego nie można wobec tego uznać za układ kontrolny i oddzielić go od systemu kontrolowanego, jak to ma miejsce w przypadku maszyn. Dennett twierdzi, że właśnie to odkrycie sprawiło największą trudność funkcjonalistom. Był to niewątpliwie jeden z ważnych powodów, które wpłynęły na osłabienie funkcjonalizmu i sprowadzenie go do funkcjonalizmu pojęciowego, stanowiącego według Kima wyjaśnienie superweniencji własności mentalnych na fizycznych[251].

Ani funkcjonalizm, ani teoria superweniencji nie mówią jednak nic na temat właściwych istocie umysłu związków *determinacji*. Tymczasem, w kwestii istoty umysłu powinno się mówić nie o oddziaływaniu przyczynowo-skutkowym struktur, lecz właśnie o ich determinacji[252]. Posługując się pochodzącym z teorii emergencji pojęciem *downward causation* („prekauzalizacji"), można sformułować termin „predeterminacji" (*downward determination*), aby następnie poszukiwać czynników predeterminujących strukturę mózgu, która z kolei determinuje, jaki typ stanu mentalnego nastąpi po innym typie stanu mentalnego. Tak, jak struktura mózgu determinuje występowanie określonych typów stanów umysłowych, tak pojawienie się poszczególnych egzemplarzy stanów umysłowych w konkretnych mózgach jest determinowane przez struktury informacyjne sygnałów docierających do mózgu z otoczenia za pośrednictwem organów sensorycznych.

[250] Na podstawie takiej analogii Fodor głosił modularną koncepcję umysłu i wpadł na pomysł języka „mentaleskiego"; według Fodora człowiek jest po prostu komputerem zaprogramowanym przez ewolucję.
[251] Można by powiedzieć, że ciężar problemów funkcjonalistów przesunął się z homunkulusa na ciało.
[252] Tak, jak ukształtowanie terenu nie oddziałuje przyczynowo na bieg rzeki, lecz go determinuje. Zawarte w dalszym ciągu propozycje zaczerpnąłem z opracowań Andrzeja Chmieleckiego.

Należy odróżniać związki typu funkcjonalnego od związków przyczynowo-skutkowych, które są jedynie czynnikami realizacji tych pierwszych. Można wówczas postulować odejście od funkcjonalizmu przyczynowo-skutkowego na rzecz funkcjonalizmu, według którego funkcje stanów umysłowych są *determinowane* przez struktury informacyjne. Teza o wielorakiej realizacji byłaby wtedy równoznaczna tezie o możliwości kodowania informacji na różnych nośnikach. Funkcjonalizm taki musiałby jednak uwzględnić fakt, że informacja jest zhierarchizowana. Utożsamianie informacji z jej nośnikiem i wynikający z tego brak pojęcia informacji różnych rzędów sprawia funkcjonalistom nie mały kłopot, a to dlatego, że informacja pojęta strukturalnie, a nie jako sygnał, jest kluczem do określenia natury umysłu i odpowiedzi na pytanie, czy maszyny mogą myśleć.

Rozdział III
Sztuczna Inteligencja

Każdy może powiedzieć o sobie, że myśli *o czymś*, że *coś* rozumie lub czegoś nie rozumie. Intuicyjnie stwierdzamy o innych ludziach to, co jesteśmy w stanie powiedzieć o sobie. Niektórzy jednak czasami powątpiewają w ludzką rozumność, upierając się przy tym, że odpowiednio zaprogramowany komputer może rozumieć coś tak samo, jak człowiek. Czynią tak, utożsamiając zazwyczaj rozumienie z myśleniem i inteligencją. Inteligencję zaś kojarzą wyłącznie z wykorzystaniem nabytej wiedzy, formułując twierdzenia na podstawie naukowo znormalizowanych pomiarów sposobu użycia zasobów intelektualnych. Jest to podejście błędne. Inteligencja pozwala nie tylko na wykorzystywanie wiedzy, ale przede wszystkim stanowi podstawowy warunek jej nabycia, decydując choćby o umiejętności rozstrzygania, czy coś jest ważne, czy nie.

Należy przy tym odróżnić bycie inteligentnym od, podlegającego naukowym pomiarom, inteligentnego zachowania. Inteligencja wiąże się z umiejętnością całościowego uchwycenia sytuacji. Opiera się ona nie tylko na dokonywaniu operacji na elementach, lecz przede wszystkim na wiązaniu ich ze sobą i umiejętnym wykorzystaniu *rozumienia* sytuacji[253]. Komputery działające szeregowo nie są w stanie „ogarnąć całości", wykonując jedynie instrukcje w określonej przez program kolejności[254]. W tym punkcie zbliżamy się już do definicji myślenia.

Myślenie bywa definiowane za pomocą pojęcia semantyki. Semantyka jest jednak wynikiem myślenia, a nie jego zasadą i ściśle wiąże się z językiem, który należy uważać za wtórne wobec myślenia narzędzie komunikacyjne. Jest to narzędzie, którego użycie wymaga myślenia. Takie założenie tkwiło u podstaw zasad testu Turinga. Czy jednak komputer rzeczywiście *używa* języka, czy też używają go tylko programiści, których polecenia maszyna wykonuje? Jak należy zdefiniować myślenie? Czy komputer myśli?

Ten rozdział poświęcony będzie sformułowaniu odpowiedzi na te pytania. Wymaga to jednak zarysowania pewnych granic dla prowadzonych rozważań. Wiąże się to z wprowadzonym przez Johna Searle'a rozróżnieniem mocnej i słabej Sztucznej Inteligencji.

3.1. Mocna i słaba Sztuczna Inteligencja

Dualizm własności fizycznych i mentalnych to współczesne, fizykalistyczne „wcielenie" problemu umysłu i ciała. Według Johna Searle'a, rozwiązaniem tego problemu może być wprowadzenie dziedziny procesów biologicznych, jako niezbędnych dla powiązania procesów fizycznych z duchowymi. Searle podąża w dobrym kierunku, ale uważa organizmy

[253] Warto tu powrócić do rozdziału pierwszego, w którym zauważyłem, że według Arystotelesa, *podstawą* myślenia jest zdolność „sprowadzania większej ilości wyobrażeń do jedności". Wyrazem inteligencji zwierzęcej jest na przykład podstęp podczas polowania.

[254] Pokładanie nadziei w komputerach działających równolegle byłoby zasadne, gdyby jakąś sytuację można było uchwycić całościowo, jako graf w pamięci operacyjnej maszyny i odnieść ją do zasobów pamięci trwałej komputera.

biologiczne za tak bardzo specyficzne, że nie można ich nazwać układami przetwarzającymi informacje. Komputery bez wątpienia są układami informacyjnymi, ale ich cyfrowa struktura jest nieporównywalnie prostsza od złożonej, analogowej struktury organizmów biologicznych. Najistotniejsze jest jednak to, że organizm biologiczny jest układem cybernetycznym, co zakłada jego autonomiczność i samodzielność.

Wbrew temu, co twierdzi Searle, organizm *jest* układem informacyjnym, ale bardzo skomplikowanym i wchodzącym w interakcje ze środowiskiem, poznającym otoczenie i konstytuującym w ten sposób własny „wirtualny świat" (kosmos wewnętrzny). Moment dostrzeżenia wagi interakcji z otoczeniem nazywany jest czasami „przełomem kognitywistycznym", który dał początek psychologii kognitywnej. Kognitywiści identyfikują umysł z komputerem cyfrowym, co okazało się ideologią inspirującą do rozpoczęcia prac nad projektowaniem maszyn interaktywnie działających w środowisku.

Jako że stany mentalne mogą pojawić się tylko w układzie informacyjnym, który jest zdolny do samodzielnego generowania informacji, a nie tylko ilościowego operowania zasobami informacyjnymi, wielu filozofów porzuciło tradycyjnie ujętą metaforę komputerową. W zamian skupiono się na symulowaniu pracy ludzkiego mózgu poprzez tworzenie sztucznych sieci neuronowych w oparciu o tezy koneksjonistyczne. Procesy zachodzące w mózgu, które według koneksjonistów można symulować obliczeniowo, są przy tym fizykalistycznie utożsamiane z procesami umysłowymi[255].

Wraz z pojawieniem się koneksjonizmu, nastąpił rozłam wśród badaczy Sztucznej Inteligencji na zwolenników i przeciwników tej koncepcji. Ci ostatni pozostali wierni tradycji komputacjonizmu. Pierwsi zaś przyjmują formułowaną przez koneksjonistów tezę o możliwości obliczeniowej symulacji funkcjonowania mózgu, jako równolegle działającego komputera cyfrowego (tak pojmują mózg kognitywiści), i jeżeli są zwolennikami funkcjonalizmu, to tylko jego słabej, „pojęciowej" odmiany. Do słabego funkcjonalizmu należą koncepcje, które nie mówią o myślących maszynach, lecz starają się zgłębić, jak myślą ludzie. To tutaj mieści się psychofunkcjonalizm i koncepcja języka „mentaleskiego" Fodora (mocno wsparta gramatyką generatywną Chomsky'ego). Tutaj też można zaklasyfikować materialistyczną teorię umysłu Armstronga, czy Kima. Nurtowi temu bliska jest słaba Sztuczna Inteligencja[256].

Komputacjoniści identyfikują mózg z komputerem, a umysł z programem komputerowym i są przedstawicielami mocnej odmiany funkcjonalizmu. Ich przekonanie o możliwości stworzenia myślących maszyn stanowi część ideologii mocnej Sztucznej Inteligencji, której korzeni należy doszukiwać się w koncepcji Alana Turinga i Hilarego Putnama.

W niniejszym rozdziale skupię się na tym, co stanowi część wspólną ideologii mocnej i słabej Sztucznej Inteligencji, a mianowicie na twierdzeniu, że właściwym poziomem opisu umysłu jest poziom składni, czyli poziom mechanicznego manipulowania symbolami zgodnie z pewnymi regułami. Od tej tezy wychodzą reprezentanci mocnej Sztucznej Inteligencji, gdy twierdzą, że komputer cyfrowy może myśleć. Tezą tą kierują się również badacze z kręgu

[255] Koneksjoniści używają komputerów cyfrowych do modelowania zachodzących w mózgu (umyśle) procesów poznawczych, ale nie są zwolennikami poglądu, że mózg (umysł) działa tak, jak standardowy komputer, czyli szeregowo. Uważają natomiast, że mózg (umysł) można porównać do komputera działającego równolegle i jako taki może być symulowany cyfrowo.

[256] Podział na mocną i słabą Sztuczną Inteligencję wprowadził John Searle, między innymi w książce *Umysł, mózg i nauka* (*op. cit.*, s. 25-26).

słabej Sztucznej Inteligencji, gdy w celu badania myślenia, tworzą komputerowe symulacje mózgu, jako systemu przetwarzania informacji. Czy rzeczywiście „maszyna składniowa" może myśleć i dzięki temu napędzać „maszynę semantyczną"?

3.2. Składnia i semantyka

Podrozdział ten można uznać za niezbędne uzupełnienie dotychczas opisanych teorii funkcjonalistycznych, które bez zawartych poniżej treści mogą wydawać się nie w pełni zrozumiałe. Mówi się czasami, że „diabeł tkwi w szczegółach" i w tym przypadku rzeczywiście można to powiedzenie odnieść do problemów, z którymi borykają się teoretycy Sztucznej Inteligencji. Podrozdział ten jest początkiem drogi ukazania tych problemów.

Pojęcia składni i semantyki ściśle wiążą się z pojęciem znaku. O znaku można rozprawiać w sposób naukowy lub filozoficzny. Najlepszym przykładem naukowego podejścia do koncepcji znaku jest *semiologia* Ferdynanda de Saussure'a, stanowiąca według tego szwajcarskiego językoznawcy część psychologii społecznej. Za spadkobierców takiego poglądu można uważać Noama Chomsky'ego i Jerry'ego Fodora. Zgodnie z tą tradycją, rozróżnić należy abstrakcyjny system językowy (*langue*, czyli po prostu to, co nazywamy *językiem*, a Chomsky nazwał *kompetencją językową idealnego użytkownika znaków*) od indywidualnych wypowiedzi ludzi posługujących się językiem (*parole*, czyli *jednostkowe wykonanie*, generowane według Chomsky'ego przez reguły językowe). Język pojęty jako *langue* to idealna struktura językowa tworząca system, w którym można wyróżnić zbiór znaków elementarnych (*alfabet*) oraz *reguły* formowania z tych znaków nieograniczonej ilości znaków złożonych, które podlegają transformacji, określonej pewnymi regułami. Sposoby łączenia i przekształcania znaków nazywa się *składnią*, *syntaktyką* lub *syntaksą*.

Składnia nic jednak nie mówi o odsyłaniu, a przecież znak to według de Saussure'a całość złożona z elementu znaczącego (*signifiant*), który odsyła do elementu znaczonego (*signifie*). Aby język był systemem znaków, musi zawierać element określający sposób interpretacji znaków, przypisywania im znaczenia. Związki znaczeniowe, czyli związki signifiant z signifie, bada semantyka. Poziom syntaktyczny i semantyczny to dwa różne poziomy, a ich relacje są tematem przedstawionych już rozważań Chomsky'ego i Fodora[257].

Warto zauważyć, że może istnieć system znaków pozbawiony składni, który nie tworzy żadnego języka. Nie jest to system twórczy, ponieważ nie posiada poziomu syntaktycznego i nie można w nim tworzyć znaków złożonych, których znaczenie zależne jest od kolejności znaków elementarnych (ustalonej regułami syntaktycznymi)[258]. System znaków bez składni jest systemem znaków umownych, czyli symboli, które odsyłają zawsze do jakichś stanów rzeczy. Twierdzenie, że syntaktyka jest konieczna dla semantyki, już w tym miejscu okazuje się zatem wątpliwe. To, że gramatyka generuje znaki złożone, które mają jakieś znaczenie, nie tłumaczy, skąd bierze się znaczenie znaków elementarnych. Co więcej, samo realizowanie reguł gramatycznych i tworzenie za ich pomocą znaków złożonych nie jest równoznaczne z „tworzeniem" znaczenia tych znaków, co ukażę w dalszej części tego rozdziału.

Tak, jak składnia nie wyprowadza poza język, tak i składnia, i semantyka abstrahują od użytkowników znaków, a przecież, jeśli mówimy o rozumieniu i interpretacji znaków, to nie

[257] Poprawność syntaktyczna nie musi oznaczać poprawności semantycznej. Zdanie „woda się pali" jest poprawne pod względem składni, ale nie jest poprawne pod względem semantycznym.
[258] Przykładem niesyntaktycznego systemu znaków jest system znaków drogowych.

możemy pomijać użytkownika, który dzięki posiadanej kompetencji językowej może znaki rozumieć i interpretować. Odniesieniem znaków do ich użytkowników zajmuje się *pragmatyka*. Rozpatrywanie kwestii związanych z myśleniem i rozumieniem powinno odbywać się na trzech poziomach: pragmatycznym (rozumienia), semantycznym (związków znaczeniowych) i syntaktycznym (reguł łączenia i przekształcania znaków)[259].

Według mnie, argument „chińskiego pokoju" Johna Searle'a można uznać za próbę ukazania, że jeśli przyjmie się, iż użycie znaków zakłada ich rozumienie i interpretację, to komputer, jako maszyna działająca tylko na poziomie syntaktyki, po prostu nie używa znaków. Komputer automatycznie wykonuje polecenia programu, a o używaniu znaków można mówić tylko w odniesieniu do programistów i użytkowników tego komputera, którzy te znaki rozumieją.

3.2.1. Semiologia, składnia i „chiński pokój" Johna Searle'a

Tak, jak Turing chciał opisać sposób pracy ludzkiego umysłu na przykładzie swojej abstrakcyjnej maszyny, tak John Searle postanowił opisać pracę uniwersalnej maszyny Turinga, czyniąc człowieka elementem jej wewnętrznego mechanizmu[260]. Człowiek ten siedzi w pokoju i ma do dyspozycji pięć plików papierów. Trzy z tych plików, nazwane „scenariuszem", „opowieścią" i „pytaniami" zawierają niezrozumiałe chińskie symbole (znaki elementarne). Dwa z tych plików, nazwane „programami", zrozumiałym dla tego człowieka językiem opisują reguły określające zasady korelowania ze sobą chińskich symboli na podstawie ich kształtów. Jeden plik opisuje korelacje symboli „opowieści" z symbolami „scenariusza", a drugi przedstawia instrukcje reagowania na kształty symboli „pytań" w odniesieniu do „opowieści" i „scenariusza" (poziom syntaktyki).

Jeśli „programy" będą perfekcyjnie napisane, a siedzący w „chińskim pokoju" człowiek (swego rodzaju *homunkulus*) bardzo sprawny w korelowaniu chińskich symboli, to Chińczyk mógłby na zadane po chińsku pytanie, uzyskać w języku chińskim syntaktycznie poprawną odpowiedź mimo tego, że siedzący w pokoju człowiek nie ma o tym języku żadnego pojęcia. Odpowiedź ta mogłaby być jednak niepoprawna semantycznie. Czy można wobec tego powiedzieć, że jeśli „chiński pokój" umie odpowiadać na zadane po chińsku pytania, to *rozumie* język chiński?

Searle twierdzi, że „chiński pokój" działa na zasadzie operowania symbolami, co pozwala mówić w tym przypadku tylko o syntaktyce, a nie o semantyce. „Chiński pokój" nie umie znaleźć związku tego, co znaczące z tym, co znaczone. Potrafi jedynie mechanicznie korelować ze sobą symbole. Nie może tu być mowy o rozumieniu, czy myśleniu, ponieważ składające się na odpowiedzi chińskie słowa nie mają dla „chińskiego pokoju" żadnego odniesienia przedmiotowego. Co więcej, dla człowieka pracującego w pokoju nie są one nawet słowami, ponieważ składają się chińskich „krzaczków", o których człowiek ten może nie wiedzieć nic – nawet tego, że są symbolami, które dla jakiegoś Chińczyka coś znaczą.

---

[259] Nie wymieniam tych poziomów w kolejności przypadkowej, gdyż uważam, że aby ktoś stał się kompetentnym użytkownikiem znaków, musi te znaki rozumieć. Najprostszym dowodem takiego twierdzenia jest to, że każdy człowiek najpierw uczy się rozumieć wypowiadane słowa, czyli chwytać ich sens. Dzieje się tak dzięki wskazaniom na przedmioty i słowom, które towarzyszą tym wskazaniom. Słowa te stają się znakami dzięki poznaniu ich odniesienia przedmiotowego. Dopiero później człowiek nabywa wiedzę na temat gramatyki i poznaje litery, z których poszczególne słowa się składają.

[260] Argument „Chińskiego pokoju" formułowany jest przez Searle'a bardzo często. Przypuszczam, że po raz pierwszy argument ten pojawił się w rozprawie *Umysły, mózgi i programy* (w: *op. cit.*, B. Chwedeńczuk (red.), s. 301).

Stąd wynika podstawowy zarzut Searle'a dotyczący behawiorystycznych podstaw testu Turinga. Searle słusznie zauważa, że podyktowane programem zachowanie maszyny, które w świecie ludzi byłoby uznane za jakąś oznakę myślenia, nie może być dowodem na to, że ta maszyna myśli, a jedynie potwierdzeniem, że jest zdolna symulować myślenie. Tak, jak nikt nie otwiera parasola, dokonując symulacji komputerowej deszczu, tak nie ma powodu sądzić, że komputer symulujący rozumienie znaków rzeczywiście je rozumie.

Wniosek taki można wyciągnąć również z wywodu poczynionego przez Hilarego Putnama w eseju *Mózgi w naczyniu*. Putnam potwierdza, że syntaktyczna gra może tylko bardzo dobrze imitować inteligentną rozmowę – tak, jak krzywa nakreślona na piasku przez mrówkę może przypominać Churchilla. Mrówka mogłaby nakreślić tą samą krzywą, nawet gdyby Churchill nigdy nie istniał. Oznacza to, że rozmowa z maszyną może być swobodnie prowadzona, podczas, gdy cały świat przedmiotów, których rozmowa ta dotyczy, znajduje się wyłącznie w umysłach osób programujących maszynę. Wynika z tego, że „test Turinga w żaden sposób nie pozwala wyeliminować maszyn zaprogramowanych *wyłącznie* do gry w naśladownictwo, i że maszyna, która nie potrafi nic *poza* uprawianiem gry w naśladownictwo, *zdecydowanie* do niczego nie odnosi swoich zagrań, podobnie jak magnetofon"[261]. Nie może tu być mowy o myśleniu.

Searle uznaje jednak, że maszyna cyfrowa może myśleć. Nie ma tu żadnej niezgodności, gdyż filozof ten przyjmuje, że człowiek mogący naturalnie ucieleśniać dowolną liczbę programów komputerowych, jest w jakiejś mierze maszyną cyfrową. Według Searle'a, myśleć może tylko maszyna wyposażona w taki system nerwowy, jak ludzki, gdyż stworzenie duplikatu przyczyn pozwala duplikować skutki. Trzeba jednak zauważyć, że Searle nie wchodzi tym w sferę sugestii konstruowania maszyn myślących, gdyż konstrukcja taka wymagałaby pełnej duplikacji procesów zachodzących w układzie nerwowym, które „stanowią część naszej biologicznej historii naturalnej"[262]. Jak zaznaczyłem w rozdziale o kłopotach z funkcjonalizmem, kauzalne wyjaśnienie umysłowości nie jest wyjaśnieniem dobrym, a Searle odnosi się w tej sprawie tylko do przyczyn i skutków. Być może z tego wynika fakt, że argumentacja tego filozofa budzi tak wiele wątpliwości, wywołując fale krytyki.

Ned Block uważa, że „najlepsza krytyka argumentu chińskiego pokoju skupiła się na tym, co Searle nazwał odpowiedzią systemową"[263]. Mówi ona, że nie możemy ze zdania „Jan nie pierze brudnych pieniędzy" wnioskować, że „Firma Jana nie pierze brudnych pieniędzy". Tak samo, nie wolno nam na podstawie czyjegoś braku znajomości chińskiego uznać, że system, którego ten ktoś jest częścią, też nie rozumie chińskiego. Neurony nie dysponują przecież indywidualnym rozumieniem, a jednak człowiek, jako swego rodzaju system, rozumie. Dlatego należy przyjąć, że „chiński pokój", jako system składający się z człowieka, plików papieru, programu oraz wejścia i wyjścia, rozumie chiński. Człowiek pełni w tym systemie rolę procesora i nie musi rozumieć chińskiego tak, jak nie rozumieją go inne komponenty systemu. Jak zauważa Block, mikroprocesory myślących komputerów same by nie myślały.

Searle w odpowiedzi na ten zarzut umieszcza cały „chiński pokój" z jego zawartością w pamięci człowieka, co ma zatrzeć analogię „chińskiego pokoju" i komputera, jednak to, że poszczególne elementy systemu nie są widoczne, nie znaczy, że przestały istnieć. Ich funkcje pełni w tym przypadku ludzki umysł i jeśli w tradycyjnym „chińskim pokoju" należało

[261] H. Putnam, *Mózgi w naczyniu*, w: H. Putnam, *op. cit.*, 1998, s. 310-311
[262] J.R. Searle, *op.cit.*, 1999, s. 15.
[263] N. Block, *op. cit.*, www.nyu.edu.

przepisywać rzeczywiste kartki papieru, to w „pamięciowym chińskim pokoju" zastępuje się je ich wyobrażeniem. Człowiek, który zapamiętał cały system „chińskiego pokoju", nadal nie rozumie języka chińskiego. Możliwe, że jego zachowanie może na to wskazywać, jednak nadal nie mamy tu do czynienia z rozumieniem chińskich symboli. Nie są one dla tego człowieka znakami, lecz tylko „zawijasami", które można potraktować tak, jak rewersy elementów układanki puzzle, których kształt ma być jedyną wskazówką do jej ułożenia[264].

Stanisław Lem twierdzi, że jeśli ktoś układa puzzle, opierając się tylko na kształcie poszczególnych fragmentów układanki, a więc działając tak, jak program komputera, wcale nie musi wiedzieć, jaki obraz z tego powstanie, ani nawet, że jakikolwiek obraz ma powstać. Według Lema, nie jest to żaden argument przeciwko Sztucznej Inteligencji. Komputer mógłby przecież wykonać dokładnie takie same kroki w procesie układania puzzle, co człowiek. Według Lema nie dowodzi to ani tego, że komputer nie rozumie swych poczynań, ani tego, że człowiek rozumie swoje[265].

Lem idzie dalej w swej kontrargumentacji i odnosi ją do siebie, czyli osoby rozumnej, ale przygłuchej. Stwierdza, że jeśli człowiek siedzący w „chińskim pokoju" pomyliłby się w identyfikacji chińskich „zawijasów" i w rezultacie dał złą odpowiedź na wyjściu, to pomyłka ta byłaby podobna do niedosłyszenia komunikatu przez Lema. A nikt przecież nie wątpi, że gdyby Lem słyszał dobrze, to by ten komunikat zrozumiał. Dlaczego więc na tej podstawie wątpić w to, że „chiński pokój" rozumie język chiński?

Searle nie zgodziłby się zapewne na to, aby obrazem ludzkiego umysłu był „chiński pokój", w którym pracuje Chińczyk, doskonale rozumiejący symbole swego języka, bo przecież żadnego homunkulusa w ludzkich głowach nie ma. Jeśli jednak by się tam znajdował, to pytanie go o rzeczy, których nie rozumie, dowiodłoby tylko, że tych właśnie rzeczy nie rozumie, a nie tego, że w ogóle nie jest zdolny do rozumienia czegokolwiek. Nie miałoby też sensu zadawanie mu pytań, na które odpowiedzi sami mu wcześniej daliśmy (na przykład w postaci pliku papierów z napisem „program"), bo dowiodłoby to tylko tego, że to my znamy odpowiedzi. Nie na tym polegał też test Turinga, który odnosił się do przejawiającego się w zachowaniu komputera *twórczego* charakteru języka. Rzeczywiście – język jest twórczy ze względu na posiadanie poziomu syntaktycznego i w tym sensie argument Searle'a nie trafia w zasadność testu Turinga. Searle nie neguje faktu, że komputer może symulować działanie „syntaktycznej maszyny", ale nie zgadza się z tym, że sama syntaktyka może zaowocować rozumieniem symboli. Searle chciał pokazać, że „składania nie jest ani konieczna, ani wystarczająca dla semantyki"[266], choć uczynił to dość niefortunnie, o czym świadczy ilość zarzutów postawionych wobec jego argumentacji[267].

Searle ma rację, gdy twierdzi, że „samo tylko rozporządzanie składnią (aspektem syntaktycznym) nie wystarcza do tego, by posiąść warstwę znaczeniową (semantykę)". Słuszne jest też twierdzenie, że „posiadanie symboli samo w sobie, albo samo tylko manipulowanie symbolami nie wystarcza do tego, by zrozumieć, co one znaczą"[268].

---

[264] Porównania takiego użył Stanisław Lem w eseju *Tajemnica chińskiego pokoju*, http://lem.arg.pl/.
[265] Jak stwierdza Lem, „ma to tyle wspólnego z «obaleniem» tezy o AI, co teza, iż z kremowych ciastek można ułożyć napis negujący szansę wybuchu Etny. Jedno za grosz nie ma nic wspólnego z drugim" – *ibidem*.
[266] J. R. Searle, *Czy intelekt mózgu jest programem komputerowym?*, w: *Świat nauki*, Nr 1, Warszawa, Lipiec 1991, s. 11.
[267] Nie będę tu tych zarzutów przytaczał, ponieważ jest ich tak wiele i są tak różnorodne, że sama ich analiza wymagałaby innej obszernej pracy.
[268] *Ibidem*.

Jak zauważyłem wcześniej, może istnieć system symboli pozbawiony składni. To, że system taki nie pozwala tworzyć języka nie znaczy, że jest systemem samodzielnym. System ten *wymaga* języka dla określenia odniesień przedmiotowych symboli należących do tego systemu. Symbole są znakami, które reprezentują poziom semantyczny dlatego, że odsyłają do czegoś na mocy *umowy*. Język przestaje być potrzebny dopiero wtedy, gdy mamy do czynienia ze znakami *nieumownymi*, o czym wspomnę jeszcze w dalszym ciągu tej pracy.

Wyposażenie pozbawionego składni systemu symboli w reguły gramatyczne pozwala tworzyć znaki złożone, na przykład słowa. Na zakończenie rozdziału poświęconego koncepcji Fodora ukazałem, że zrozumienie jakiegoś słowa jest możliwe mimo braku w nim niektórych symboli. Nie dzieje się tak ze względu na syntaktykę, choć rzeczywiście, zastosowanie stosownych reguł pozwoliło zbudować z poszczególnych symboli określone słowo, a znajomość tych reguł pomaga w zrozumieniu tego słowa. Od chwili powstania słowo to zaczyna jednak „żyć własnym życiem", stając się jakby nowym znakiem (złożonym z poszczególnych symboli).

Słowo „drzewo" jest połączone znaczeniem ze wszystkimi znakami umownymi, które przywołują nam na myśl *drzewo*. Jeśli w słowie tym występują jakieś braki (np. „dr*ewo"), to jest ono nadal dla nas zrozumiałe nie tylko ze względu na fakt, że znamy syntaktyczną rolę symboli, które to słowo tworzą, ale przede wszystkim dlatego, że jego znaczenie poznaliśmy już wcześniej.

Przypuśćmy, że znamy język chiński. Gdybyśmy zobaczyli słowo „dr*e*o", a obok ktoś pokazałby nam chiński symbol drzewa, to bez większego problemu zgadlibyśmy, co „wybrakowane" słowo znaczy (chyba, że kształt chińskiego symbolu również wzbudzałby nasze wątpliwości). Ktoś mógłby być na tyle złośliwy, że kazałby nam zgadnąć, co może znaczyć słowo „dr***o". Moglibyśmy przypuszczać, że ma ono to samo znaczenie, co słowo „drzewo", ale równie dobrze mogłoby mieć znaczenie słowa „drewno". W rozwiązaniu zagadki nie pomoże tu znajomość syntaktycznie określonej roli symboli, bo zbyt wiele symboli jest ukrytych, a przecież znaczenie obu słów ciągle jest nam znane. Same reguły operowania symbolami nie determinują znaczenia słów, które za pomocą tych reguł zostały utworzone z poszczególnych symboli. I choć symbole jako takie zawsze do czegoś odsyłają (inaczej nie byłyby znakami), to ich interpretacja wymaga *wiedzy*, do czego.

Jak już wcześniej zauważyłem, argument „chińskiego pokoju" można pojmować jako sprzeciw wobec behawiorystycznego rodowodu testu Turinga. Twierdzenie, że inteligentne zachowanie jest równoznaczne z posiadaniem intelektu, Searle uznaje za spadek po usilnych próbach uczynienia z psychologii nauki ścisłej. W mocnej Sztucznej Inteligencji filozof ten dopatruje się także dualizmu, opartego na uznaniu umysłu za niezależny od mózgu program komputerowy. Wyraża przy tym odważną tezę, że „rozum jest takim samym zjawiskiem biologicznym, jak trawienie"[269]. Ważną konkluzją, jaka wynika z argumentu „chińskiego pokoju" jest spostrzeżenie, że symulacja nie jest duplikacją, a symulować obliczeniowo można wszystko.

Przedstawiciele słabej Sztucznej Inteligencji porównują ludzki mózg (i zarazem umysł) nie do samego programu komputerowego, lecz do komputera cyfrowego. Idąc śladami Chomsky'ego i Fodora zauważają, że reguły gramatyczne, których przestrzegają użytkownicy języka, są podobne do tych, którymi kieruje się komputer. Jak stwierdza Searle, komputery nie

[269] *Ibidem*.

przestrzegają żadnych reguł. Ich automatyczne działanie jest podyktowane formalnymi procedurami, które tylko w oczach człowieka są regułami. Komputer ich nie przestrzega, tylko działa tak, *jak gdyby* ich przestrzegał. Problem polega więc na dosłownym potraktowaniu pewnej wygodnej metafory.

Searle twierdzi, że „komputer prawdopodobnie nie jest ani lepszą, ani gorszą metaforą mózgu, niż inne mechaniczne metafory", a było ich wiele[270]. W zależności od osiągnięć naukowych, mózg i umysł były porównywane do młyna, silnika parowego, systemu telegraficznego, czy centrali telefonicznej. Nie dziwi więc Searle'a fakt, że współcześnie metaforą mózgu jest komputer.

3.2.2. Semiotyka, myśl i teoria znaku Charlesa S. Peirce'a

Charles Sanders Peirce (1839-1914) najbardziej znany jest jako twórca pragmatyzmu, jednak w niniejszej pracy zajmę się tylko podstawą zasady pragmatycznej, ujętą w *semiotyce*, czyli ogólnej teorii znaku.

Peirce zgodził się z Kartezjuszem, że podstawą poznania nie mogą być dane zmysłowe. Uważał bowiem, że postrzeżenie nie jest aktem czysto zmysłowym, lecz wiąże się z aktywnością umysłu, jest rozumieniem, które wymaga wyjścia poza dane zmysłowe. Tu jednak zaczyna się polemika z kartezjanizmem. Peirce stwierdził, że w poznaniu nie można zaczynać od wątpienia, lecz od przekonań (uznawanych sądów bądź myśli). Według Peirce'a, poznanie po prostu składa się z przekonań. Jest ono *zmediatyzowane* (zapośredniczone) i stanowi uobecnienie czegoś za pomocą czegoś innego, czyli znaków, utożsamianych przez Peirce'a z reprezentacjami[271].

Zasadniczą właściwością znaków jest to, że informują one o czymś w sposób zastępczy, przy czym nigdy nie ma się do czynienia z bezpośrednią obecnością tego, co znak reprezentuje. Reprezentacja umożliwia więc tylko pośredni kontakt z rzeczywistością i tylko z tym mamy do czynienia w poznaniu. Zgodnie z koncepcją Peirce'a, myśl jest znakiem, ponieważ zawsze jest *o czymś*, jest czymś uchwyconym jako reprezentacja *czegoś*. W tym ujęciu znak jest jedynym sposobem istnienia myśli.

Ważne jest to, że Peirce rozumiał myśli obiektywistycznie, abstrahując od podmiotu myślącego. Uniezależniał tym samym myśli od mózgu i pojmował je w sposób logistyczny, jako ogniwa w niezależnych od podmiotu łańcuchach przesłanek i wniosków. Myśl, jako reprezentacja, jest według Peirce'a znakiem przekładalnym na inne znaki, co oznacza, że poznanie nie może mieć punktu początkowego (można bez końca cofać się do logicznych przesłanek, z których żadna nie jest oczywista). Nie jest wobec tego możliwe poznanie bez

[270] Warto tu przytoczyć obszerniejszy cytat: „Załóżmy, że ktoś wie, jak działa zegar. Powiedzmy, że okropnie trudno wyobrazić sobie jak działa zegar, bo chociaż jest ich wiele wokół, nikt nie wie, jak zbudować zegar, a wszelkie próby zbadania ich pracy prowadzą do zniszczenia zegarów. Przypuśćmy teraz, że grupa naukowców powiada: «Będziemy rozumieć jak działa zegar, jeśli zbudujemy maszynę funkcjonalnie mu równoważną, maszynę, która będzie mierzyć czas tak dobrze, jak zegar». Zbudowali oni następnie klepsydrę i stwierdzili: «Teraz wiemy jak działa zegar», albo: «Jeśli tylko zbudujemy klepsydrę tak dokładną jak zegar, będziemy końcu wiedzieć, jak on działa». Jeśli w tym porównaniu zamienimy «zegar» na «mózg», a «klepsydrę» na «program komputera cyfrowego», zaś «mierzenie czasu» zamienimy na «inteligencję», będziemy mieli opis całej niemal sytuacji w badaniach nad sztuczną inteligencją i w naukach o poznawaniu." – J. R. Searle, *op. cit.*, 1995, s. 50-51.

[271] W koncepcji Peirce'a można dostrzec wpływ teorii poznania sformułowanej przez Johna Locke'a.

uprzedniego założenia jakichś przesłanek, z których każda jest podatna na falsyfikację. Peirce stwierdza, że rzeczywistość pozaznakowa jest bez pośrednictwa znaków niepoznawalna.

Podstawową kategorią semiotyki Peirce'a jest *triada*, czyli relacja trzech elementów, z których każdy jest łącznikiem pomiędzy dwoma pozostałymi. Istotą triady jest *mediacja* (zapośredniczenie) dwóch elementów przez trzeci. Według Peirce'a, znak jest taką właśnie mediująca reprezentacją. Jako trójczłonowa relacja, znak łączy ze sobą nośnik znaczenia (*reprezentamen*, *relat*, nośnik przekazu), przedmiot znaku (*korelat*) i znaczenie (*interpretant*), które jest kolejnym znakiem[272].

Peirce twierdził, że celem *semiozy* jest poznanie, co oznacza, że dzięki procesom znakowym jest możliwe zbliżenie się do prawdy. Poznanie zawsze zapośredniczone jest przez reprezentację, co nie zaprzecza realizmowi i klasycznej koncepcji prawdy. Zgodnie z teorią Peirce'a, przedmiotem znaku może być wprawdzie tylko statyczny przedmiot bezpośredni, ale poprzez kategorię tego przedmiotu, znak wskazuje na *realny* przedmiot dynamiczny.

Peirce dokonał klasyfikacji znaków, podając triadyczno-trychotomiczną strukturę znaku. Przedstawił przy tym trzy trychotomie: z uwagi na nośnik znaczenia (reprezentamen), z uwagi na relację nośnika znaczenia do przedmiotu znaku (reprezentamen-korelat) i z uwagi na relację znaku do znaczenia (znak-interpretant). Dla potrzeb tej pracy warto zwrócić uwagę na trychotomię drugą, ponieważ precyzuje ona pojęcia szeroko stosowane między innymi przez teoretyków Sztucznej Inteligencji.

Jak zaznaczyłem wyżej, druga trychotomia rozróżnia znaki z uwagi na powiązanie z ich przedmiotem, i tak:

- jeśli coś funkcjonuje jako znak dzięki własnym cechom, które przypominają przedmiot znaku, czyli na mocy strukturalnego podobieństwa, to mamy do czynienia z *ikoną*;
- jeśli coś funkcjonuje jako znak wskutek naturalnego, rzeczowego czy przyczynowo-skutkowego związku z przedmiotem, to mamy do czynienia z *oznaką*;
- jeśli coś funkcjonuje jako znak ze względu na jakąś zasadę opartą na konwencji, umowie, czy regułach, to mamy do czynienia z *symbolem*.

Można teraz przypomnieć tezę Peirce'a, że każda myśl jest ze swej natury znakiem. Znaki jednak można podzielić na dziesięć klas, opierając się na ich triadyczne-trychotomicznej strukturze. Spośród wyodrębnionych klas znaków tylko jedną uznał Peirce za klasę znaków autentycznych, które nie uległy „degeneracji", ponieważ wszystkie ich elementy mają charakter myśli. Znaki takie to (według nazewnictwa Peirce'a) *legisigny symboliczno-argumentowe*.

Nośnikiem znaczenia w *legisignie* jest coś niespostrzegalnego, „myślowego", czyli na przykład jakieś prawo, działające jako znak (można tu wymienić abstrakcyjne oddziaływanie reguł gramatycznych, co odsyła nas do treści poprzednich rozdziałówi tematu *syntaktyki*). Zdefiniowany wyżej symbol odnosi się do przedstawionych już rozważań na temat *semantyki* znaków elementarnych. *Argument* to znak, którego kontekst interpretacyjny jest pełny, a jego koniecznym i ostatecznym przedmiotem jest prawo (przykładem takiego znaku jest

[272] Należy pamiętać, że gdy mówi się tu o przedmiocie znaku, to chodzi jedynie o obiekt, którego znak dotyczy i nie ważne jest, czy obiekt ten rzeczywiście istnieje. Peirce rozróżnia przedmiot bezpośredni (statyczny element znaku, który nie oddziałuje na użytkownika) i przedmiot realny (dynamiczny, mogący oddziaływać na użytkownika znaków).

wnioskowanie, co wiąże się na przykład z omówioną koncepcją Fodora). Znak, który spełnia funkcje wszystkich wyżej wymienionych znaków, jest „czystą" *myślą*.

Według Peirce'a, poznać można tylko myśl ujętą w znaku, a myśl niepoznawalna nie istnieje. Dlatego właśnie każda myśl jest ze swej natury znakiem. Interpretacja znaku polega na przekładzie go na inne znaki, z czego wynika, że znaczeniem znaku jest zawsze inny znak. Myśl jest wobec tego znakiem będącym znaczeniem innego znaku (innej myśli) i zakłada zarazem przynajmniej jeszcze jedną myśl (znak), która jest jej znaczeniem. Taki nieskończony proces pozwala ujmować różne aspekty (znaczenia) tego samego przedmiotu, który może być symbolizowany na różne sposoby (na przykład za pomocą łacińskich liter, albo chińskich „zawijasów").

Jak zauważa Andrzej Chmielecki[273], Peirce abstrahuje od umysłu jakiegoś użytkownika języka, ale oczywistym jest, że proces semiozy musi zachodzić w umyśle. Peirce, formułując ogólną teorię znaku, nie tworzył teorii umysłu, ale wyraźnie ujmuje umysł w kategoriach logicznych.

Zgodnie z koncepcją Peirce'a, każdy interpretant (znak interpretujący inny znak) charakteryzuje się trzema stopniami przekładu. Sama *możliwość* przekładu na inne znaki to interpretant *bezpośredni*. Jeśli interpretant wywołuje w umyśle jakiś realny efekt i mamy do czynienia z jednostkowym aktem interpretacji, to takie znaczenie (pojęte w sensie psychologicznym) Peirce nazywa interpretantem *dynamicznym*. Pełne znaczenie uzyskuje się jednak dopiero na poziomie trzecim przekładu, w postaci interpretanta *normalnego*, który ma charakter czysto logiczny i jest prawem, czy regułą decydującą o efekcie, jaki wywołany by był w umyśle przez znak, gdyby ktoś ujął w swych myślach całą tę regułę.

Interpretant normalny nie wyprowadza jednak poza znaki, ku rzeczywistości. Interpretantem *ostatecznym* jest według Peirce'a *zasada pragmatyczna*, czyli znaczenie będące praktyczną konsekwencją znaku. Chodzi o takie znaczenie znaku, które jest logicznym sądem warunkowym, którego następnik jest imperatywem działania[274].

Powyższe przedstawienie semiotyki Peirce'a jest niezbędne w poszukiwaniach odpowiedzi na pytanie, czym jest myślenie i rozumienie. Powiązanie teorii znaków z koncepcją informacji pozwoli przedstawić propozycje odpowiedzi nie tylko na powyższe pytanie, ale także ustosunkować się do stanowiska zwolenników Sztucznej Inteligencji.

[273] Opieram się tutaj na Andrzeja Chmieleckiego wykładach z semiotyki.
[274] Użyteczność odnosi się więc w pragmatyzmie Peirce'a do *sensowności*, a nie do prawdziwości. W takim ujęciu zachowana zostaje tradycyjna definicja prawdy, jako zgodności sądu z rzeczywistością.

Część III
Inteligencja naturalna

Rozdział IV
Od informacji do myślenia

Aby ująć związek *znaku*, *informacji*, *rozumienia* i *myślenia*, posłużę się konceptualizacją Andrzeja Chmieleckiego. Rozpocznę od ważnej kwestii: co należy rozumieć pod pojęciem *umysłu*? Andrzej Chmielecki zauważa, że problem w udzieleniu odpowiedzi na to pytanie tkwi w bardzo szerokim pojęciu umysłu, pochodzącym z anglojęzycznej filozofii analitycznej. Od rzeczownika *mind* pochodzi przymiotnik *mental*, traktowany jako przeciwieństwo przymiotnika *physical*. Takie „pojęcie umysłu pokrywa zakresowo bardzo szeroką dziedzinę – wrażenia zmysłowe, spostrzeżenia, doznania bólu, emocje, wyobrażenia, idee, myśli, akty woli”[275]. Byty, które obejmuje pojęcie *mind*, różnią się jednak od siebie typem prawidłowości, którym podlegają, a więc mają odmienną *naturę*. Stany umysłowe nie są po prostu stanami nie-fizycznymi. Można je raczej określić mianem *stanów ducha*. Pomiędzy dziedziną stanów i procesów umysłowych a dziedziną stanów i procesów fizycznych należy uwzględniać poziom stanów *animalnych* oraz poziom stanów *informacyjnych*.

Według Chmieleckiego, na strukturę rzeczywistości składa się pięć warstw o różnej naturze, czyli różnym sposobie powstawania i różnych prawidłowościach funkcjonowania. Całą rzeczywistość można opisać jako składająca się z pięciu dziedzin bytowych, które tworzą poszczególne warstwy: bytów fizycznych; organizmów żywych (roślin i prostych organizmów zwierzęcych); bytów psychicznych (zwierząt posiadających animę[276]); bytów duchowych (osób) oraz bytów inteligibilnych. Podział ten dokonany został z uwagi na naturę bytów, czyli ze względu na kolektywny zbiór tych ich cech, które stanowią o przynależności bytów do danej dziedziny (cechy te określają sposób funkcjonowania bytu i jego interakcji z otoczeniem)[277].

Każda z tych warstw zakłada warstwę niższą, jako warunek swego powstania, czyli swój fundament bytowy. Za ostateczny fundament bytowy, coś bezwzględnie samoistnego, należy uznać jakąś postać bytu fizycznego, która być może dopiero zostanie odkryta. Warstwowe zróżnicowanie rzeczywistości świadczy o jej niejednorodności, natomiast relacja ufundowania o jedności.

Zaistnienie każdego bytu zależy od jego czynników realizacji (*ufundowania* bytowego), natomiast to, *czym* ten byt jest (jakie są jego cechy istotowe), zależy od czynników determinacji tego bytu (jego *umocowania* bytowego).

Wspólnym elementem wszystkich czynników determinacji bytów o takiej samej naturze jest *zasada determinacyjna*. Są cztery takie zasady: *forma* w dziedzinie bytów fizycznych,

[275] A. Chmielecki, *Intuicja intelektualna, fenomen rozumienia*, http://kognitywistyka.net/artykuly/ach_intuicja_intelektualna.pdf.
[276] W terminologii A. Chmieleckiego *anima* to psychika, która steruje funkcjonowaniem organizmu zwierzęcego.
[277] Zbiór kolektywny to zbiór *mereologiczny*, czyli całość powiązanych ze sobą części, ich struktura.

informacja w dziedzinie organizmów żywych, *reprezentacja* w dziedzinie bytów psychicznych oraz *sens* w dziedzinie bytów duchowych. Dziedzina bytów inteligibilnych jest ufundowana w dziedzinie bytów duchowych i nie ma własnej zasady determinacyjnej[278].

Z uwagi na sposób istnienia, byty można podzielić na *realne* (zdolne do działania, czyli dokonywania zmian i podatne na zmiany spowodowane działaniem z zewnątrz)[279], *idealne* (będące pod względem działania przeciwieństwem bytów realnych)[280], *irrealne* (podatne na zmiany, ale nie mające zdolności działania)[281] i *nadrealne* (mają zdolność działania, lecz nie są podatne na zmiany z zewnątrz)[282].

Egzystencjalnie najsilniejszym sposobem istnienia jest sposób istnienia bytów realnych, które są najbardziej *samoistne*. Najbardziej *samodzielne* są natomiast byty idealne. Każdy byt ma swoje miejsce w porządku ufundowania (określającym *naturę* bytu) i umocowania (określającym *istotę* bytu). Najlepszym opisem każdego bytu jest ustalenie jego miejsca w obu tych porządkach.

4.1. Informacja

W języku potocznym „informacja" jest zwykle kojarzona z wiadomością, czyli komunikatem. Komunikat to ciąg znaków nadanych przez nadawcę i/lub odebranych przez odbiorcę komunikatu[283]. Jeśli ciąg znaków został nadany celowo, jako znaczący dla ich nadawcy, to niezależnie od tego, czy nastąpił odbiór, mamy do czynienia z przekazem *komunikatywnym*. Jeśli natomiast ciąg znaków został odebrany jako znaczący dla ich odbiorcy, to niezależnie od tego, czy ciąg ten został nadany, mamy do czynienia z przekazem *informatywnym*.

O procesie *komunikacji* mówimy wtedy, gdy jakieś systemy oddziałują na siebie za pośrednictwem znaków, co ma miejsce, gdy przekazy komunikatywne nadawców są dla odbiorców przekazami informatywnymi. Użytkownicy znaków mogą komunikować się między sobą za pomocą różnych *środków przekazu*, ale z zastosowaniem tylko tych *kodów*, które są znane zarówno nadawcy, jak i odbiorcy komunikatu[284]. Spełnienie wymienionych warunków nie zapewnia jeszcze porozumienia, które jest możliwe pod warunkiem braku zaburzeń w środkach przekazu i zgodnej interpretacji przekazów informatywnych i komunikatywnych.

Znaki mogą pełnić w procesie komunikacji różne role. Karl Bühler wymienił trzy funkcje komunikacyjne znaków. W odniesieniu do nadawcy komunikatu znaki pełnią funkcję

[278] Można chyba powiedzieć, że sens jest *czynnikiem predeterminacji* bytów inteligibilnych.
[279] Bytami *realnymi* są wszystkie byty *materialne* (na przykład człowiek, jako istota cielesna).
[280] Bytami *idealnymi* są obiekty matematyczne, wartości, prawa i struktury logiczne.
[281] Bytami *irrealnymi* są przedmioty spostrzeżeń (fenomeny), czas i przestrzeń, albo takie zjawiska fizyczne, jak na przykład cień. Najważniejszymi dla tematyki tej pracy są jednak takie niefizykalne byty irrealne, jak *znaczenie* i *informacja*.
[282] Bytem *nadrealnym* jest człowiek rozumiany jako *osoba*. Dla osoby wierzącej bytem nadrealnym jest przede wszystkim Bóg pojęty jako Stwórca. Dla ateisty, który uważa Boga za mityczną postać literacką, Stwórca jest bytem irrealnym.
[283] Ciąg znaków, które zostały nadane, a nie zostały odebrane, jest komunikatem tak samo, jak może nim być dla kogoś ciąg znaków, których nikt nie nadał.
[284] Kod to sposób interpretacji znaków, które mogą tworzyć ciąg będący komunikatem.

ekspresyjną i jako takie zostały przez Bühlera nazwane *oznakami*[285]. W odniesieniu do odbiorcy komunikatu znaki pełnią funkcję *apelującą* i jako takie są *sygnałami*. W odniesieniu do rzeczywistości znaki pełnią funkcję *przedstawiającą* i jako takie są *symbolami*[286]. Komunikat może pełnić wszystkie wymienione funkcje.

Utożsamienie informacji z komunikatem pozwala uznać znaki za logicznie od niej wcześniejsze. W takim ujęciu znaki, jako nośniki informacji, mogą istnieć również w komputerze, co pozwala niektórym twierdzić, że w maszynach tych możemy mieć do czynienia z syntaksą i semantyką.

Antropomorfizowanie pojęcia informacji nie kłóci się z powszechnie przyjętym naukowym ujęciem tego terminu, pochodzącym od Claude'a Shannona. Ten amerykański matematyk prowadził rozważania czysto techniczne, na temat ilości i wierności przekazywanych komunikatów. W rozważaniach tych nawiązał o twierdzeń z dziedziny fizyki. Jedno z nich mówi, że we wszystkich układach fizycznych, które nie są pod wpływem żadnych zewnętrznych oddziaływań, zanikają wszelkie różnice (temperatury, ciśnienia, prędkości, itp.). Stan zróżnicowania jest wobec tego mniej prawdopodobny, niż stan niezróżnicowania, określany wielkością zwaną *entropią,* która w układach odosobnionych może tylko wzrastać.

Shannon potraktował entropię, jako ilościową miarę niezróżnicowania (*nieokreśloności*) stanu rzeczy, scharakteryzowanego rozkładem prawdopodobieństw. Nieokreśloność ta, czyli niepewność wynikająca z braku wiedzy o danym stanie, maleje wraz ze wzrostem prawdopodobieństwa tego stanu. Pojęcie ilości informacji Shannon powiązał właśnie z tym prawdopodobieństwem, określając ją jako funkcję prawdopodobieństwa pojawienia się określonego stanu rzeczy spośród wielu możliwych ($I = -\log_2 p$).

Określenie jakiegoś parametru jest równoznaczne z zanikiem nieokreśloności, a miarą zmniejszania się niepewności jest, według Shannona, ilość uzyskanej informacji. Takie funkcjonalne pojęcie informacji jest zgodne z tym, co potocznie się o niej mówi, że redukuje niepewność i poszerza wiedzę. Shannon po prostu opisał, *jak* to się dzieje. Wprowadził też stosowaną do dziś jednostkę miary ilości informacji – *bit* (*binary digit*). Jeden bit informacji jest równoznaczny z *wiedzą*, że pojawiła się jedna z dwóch tak samo prawdopodobnych ewentualności ($i = -\log_2 ½$).

W komputerach używa się dziś kodu binarnego, gdyż ich konstrukcja umożliwia wystąpienie tylko dwóch równie prawdopodobnych stanów, które zwykło się symbolizować dwiema cyframi: 0 i 1 (jest to cały *alfabet* komputera). Komputery są w stanie wykryć za jednym razem (w jednym takcie) na przykład 64 bity informacji. O jakości działania komputera decyduje między innymi ilość bitów wykrytych w jednym takcie oraz szybkość taktowania.

Na podstawie ilościowego ujęcia informacji można określić, ile bitów daje nam jakaś czynność, co jest przydatne głównie dla celów technicznych. Ujęcie to nazywane jest często *syntaktycznym*, ponieważ dotyczy przekazywania komunikatów, które są ciągami znaków. Komunikaty te mają jednak jakąś *treść*, a o tym Shannon nie wspomina.

[285] Oznaki są tu symptomami stanów wewnętrznych nadawcy komunikatu. Jest to jakby rozszerzenie definicji oznaki sformułowanej wcześniej w oparciu o teorię Peirce'a, który definiował oznakę jako znak związany ze swym przedmiotem w sposób naturalny. Definicja Bühlera odnosi oznakę do jego użytkownika, a konkretnie do nadawcy ciągu znaków (komunikatu), natomiast Peirce abstrahował od użytkowników znaków.

[286] U Bühlera symbol jest po prostu znakiem przedstawiającym rzeczywistość. Peirce określił mianem symbolu tylko te znaki, których relacje z przedmiotem znaku są uregulowane umownie.

Innym ilościowym ujęciem informacji jest koncepcja przedstawiona przez francuskiego fizyka, Leona Brillouin. Wziął on pod uwagę ilość możliwych stanów jakiegoś obiektu oraz ilość tych stanów, które już zostały zrealizowane. Poznanie tych stanów daje nam wiedzę dotyczącą *struktury* ilościowej tego obiektu. Brillouin uznał, że tak, jak entropia jest miarą niezróżnicowania, tak miarą zróżnicowania jest *negentropia*, a samo zróżnicowanie mierzy się w ilości bitów informacji *strukturalnej*.

Biorąc pod uwagę możliwości komputera 64-bitowego, można obliczyć, że jest on w stanie wygenerować 18 446 744 073 709 600 000 różnych znaków. Każdy z tych znaków da się opisać w postaci ciągów zer i jedynek, symbolizujących stany generowane przez maszynę w jednym takcie. Pierwszy z poniższych ciągów może na przykład być opisem znaku *A*, a drugi znaku *C*.

0001111111111000	0001111111111000
0001110000111000	0001110000000000
0001111111111000	0001110000000000
0001110000111000	0001111111111000

Oba ciągi zer i jedynek są identyczne pod względem zawartej w nich informacji ilościowej, ale nikt nie zaprzeczy, że istotnie się od siebie różnią.

Inaczej do pojęcia informacji podszedł Andrzej Chmielecki. Filozof ten stwierdza, że informacja to „wszelka zarejestrowana (wykryta) różnica stanów (parametrów) fizycznych”[287], a ponieważ zdolność rozróżniania stanów fizycznych jest podstawową cechą organizmów żywych, przeto organizmy te są *układami informacyjnymi*, czyli takimi układami, które są zdolne do wykrycia, przetworzenia i wykorzystania informacji. Stany takich układów są przy tym zależne od tego, *jaką* informację zarejestrują. Informacja zarejestrowana jest *kodem*. Informacja, jako zbiór różnic, może być kodowana w rozmaity sposób i na różnych nośnikach[288]. Bez nośnika i kodu informacja zaistnieć nie może. Jest ona bowiem abstrakcyjną formą, która wymaga materialnego „ucieleśnienia”.

Jeśli wykrycie różnicy polega na odniesieniu aktualnego stanu rzeczy do innych możliwych stanów, to mamy do czynienia z różnicą typu *selekcji* (na zasadzie ekskluzji). Jeśli natomiast wykrycie różnicy wynika z odniesienia do siebie wielu współwystępujących stanów, to mamy do czynienia z różnicą typy *zestawu* (na zasadzie koniunkcji). Wykrycie różnicy typu selekcji jest równoznaczne z uzyskaniem informacji jakościowej, natomiast wykrycie różnicy typu zestawu oznacza uzyskanie informacji strukturalnej, która składa się z informacji jakościowych.

Andrzej Chmielecki zauważa w związku z tym, że powyższy podział wskazuje na występowanie informacji różnego rzędu. Kod, czyli informację zarejestrowaną na jakimś nośniku, można uznać za informację rzędu zerowego. Informacja jakościowa, jako elementarna, jest informacją pierwszego rzędu, zaś informacja strukturalna, jako zbiór informacji jakościowych, to informacja rzędu drugiego. Informacja trzeciego rzędu jest zbiorem informacji strukturalnych, natomiast informacje wyższych rzędów to pewne struktury relacyjne, także zbudowane z informacji strukturalnych. Informacja jest więc *zhierarchizowana*.

[287] A. Chmielecki, *op. cit.*, 1999, s. 282.
[288] Można się tu dopatrywać nowego sformułowania funkcjonalistycznej tezy o wielorakiej realizacji. Jak się jednak okaże, sposób realizacji informacji nie zawsze jest bez znaczenia.

Odnosząc powyższe do przedstawionych ciągów stanów (reprezentujących znaki *A* i *C*) można powiedzieć, że komputer jest układem informacyjnym, działającym na podstawie informacji strukturalnej. Działaniem tej maszyny kierują bowiem zarejestrowane przez nią w każdym takcie poszczególne ciągi stanów[289].

Najprostsze organizmy żywe (na przykład rośliny) są układami informacyjnymi wykorzystującymi jedynie informację jakościową; z wykorzystaniem informacji strukturalnej mamy dopiero do czynienia na poziomie bytu psychicznego. Przejście od dziedziny bytów fizycznych do dziedziny bytów psychicznych polega na tworzeniu informacji wyższego rzędu poprzez uchwycenie informacji strukturalnej za pomocą receptorów. Według Andrzeja Chmieleckiego, odbywa się to w taki sposób, że działającym na receptory bodźcom fizycznym są przyporządkowywane stany centralnego układu nerwowego. Przyporządkowanie to następuje na zasadzie *morfizmów* (relacje między bodźcami znajdują swe odpowiedniki w relacjach między stanami centralnego układu nerwowego). W wyniku takiego odwzorowania, w centralnym układzie nerwowym powstają *reprezentacje* świata zewnętrznego. Reprezentacji psychicznych nie należy identyfikować ze stanami mózgu, które są jedynie sposobem kodowania informacji (informacji wyższego rzędu nie można identyfikować z informacją rzędu zerowego).

Centralny układ nerwowy może wykorzystywać informację strukturalną w celu kierowania funkcjonowaniem organizmu. Aby było to możliwe, zarejestrowane różnice stanów fizycznych (tworzywo reprezentacji psychicznych) muszą zostać odpowiednio zinterpretowane, jako informacja *o czymś*. Animalna („zwierzęca") interpretacja informacji[290] może polegać na przekładzie tego, co nieznane na to, co znane dla organizmu, poprzez uchwycenie pewnych *korelacji* (na przykład współzależności wzrokowo-motorycznych) lub *asocjacji* (występowania czegoś po czymś). Ustalone zostaje w ten sposób *znaczenie* informacji, ale nie to, czy jest to informacja *ważna* dla organizmu. Określenie wagi informacji wiąże się ze stanem układu popędowego zwierzęcia.

Wspomniany wyżej proces przekształceń typu morfizmów odbywa się równolegle do procesów neurofizjologicznych, zachodzących w mózgu, który funkcjonuje według własnego czasu systemowego. Dlatego czas i przestrzeń, jako własności fizykalne, nie są w neurofizjologii istotne. Mózg działa, taktując według swoich własnych rytmów. Nie jest przy tym ważny substrat fizyczny mózgu, lecz liczą się tylko parametry topologiczne komórek mózgowych, ich wzajemne relacje i funkcje. Mózg jest wielopoziomową strukturą neuronów, jednak jego istotą nie jest to, z czego się fizycznie składa, lecz to, że jest odpowiednio zorganizowanym, abstrakcyjnym układem funkcjonalnym. Nie interesuje nas wobec tego ani to, z czego zbudowane są neurony, ani to, co się w nich dzieje, a jedynie aktywność neuronów na ich synapsach, czyli wejściach i wyjściach.

Aktywność neuronów można scharakteryzować matematycznie, za pomocą funkcji zmiennej czasowej, czyli zmiennych w czasie parametrów ilościowych[291]. Funkcje zmiennej czasowej

[289] Słowo *AC* mogłoby przecież być jakąś komendą dla komputera, która przełożona na język binarny, odzwierciedlałaby sekwencję równoczesnego wystąpienia określonych stanów. Stany te, zgodnie z programem maszyny, mogłyby wtedy powodować przejście maszyny do innych stanów i „wyprodukowanie" innych słów.
[290] Operuję tu terminologią zaproponowaną przez Andrzeja Chmieleckiego.
[291] Chodzi o takie parametry, jak na przykład potencjał i ilość impulsów elektrycznych generowanych w jednostce czasu.

neuronów należących do danej populacji, Chmielecki nazywa *sygnałami*[292], zaś zbiór wszystkich *możliwych* sygnałów, określa mianem ich *przestrzeni*.

Jako że każdy neuron może wygenerować różne impulsy, w zależności od tego, jak zostanie pobudzony, przestrzenie sygnałów można uznać za zbiory wszystkich możliwych pobudzeń neuronów należących do danej populacji. Aktualny stan pobudzenia wszystkich neuronów tej populacji to jeden z możliwych stanów przestrzeni sygnałów (zarejestrowanie tego stanu daje zbiór informacji jakościowych).

W jednej chwili nie może zatem zostać zrealizowana cała przestrzeń sygnałów, lecz tylko pewien jej podzbiór, czyli pewna *podprzestrzeń*. Jeśli należące do tego zbioru stany przestrzeni sygnałów zostaną zarejestrowane i jeśli traktowane są integralnie jako stanowiące pewną całość, to mamy do czynienia z informacja strukturalną, którą w tym ujęciu Chmielecki definiuje, jako „wygenerowaną w jakimś układzie podprzestrzeń przestrzeni sygnałów”[293].

Biorąc pod uwagę powyższe, mogłoby się wydawać, że mocny funkcjonalizm jest teorią, którą można uznać za najbardziej adekwatną dla rozważań związanych z działaniem mózgu. Andrzej Chmielecki zwraca jednak uwagę, że teoria ta nie uwzględnia hierarchiczności informacji, pomijając istotne dla funkcjonowania mózgu informacje wyższych rzędów.

Funkcjonalista zawsze abstrahuje od sposobu realizacji i zapewne tak samo postąpiłby w sprawie informacji. Musimy jednak wziąć pod uwagę fakt, że coś, co na danym poziomie uważamy za informację, może stanowić *tworzywo* informacji wyższego rzędu i jako takie nie powinno być uznawane za nieważne. Funkcjonalistyczna teza o wielorakiej realizacji nie może uwzględniać hierarchii informacji, ponieważ sugerowałaby to, że tworzywo informacji wyższego rzędu nie ma znaczenia, a przecież bez kodu i informacji jakościowej nie byłoby informacji strukturalnej.

Sposób realizacji informacji strukturalnej jest ważny ze względu na konieczność odpowiedniej realizacji tworzywa tej informacji, czyli informacji jakościowej. Należy pamiętać, że tworzywem (czynnikiem realizacji) reprezentacji nie jest jej zapis na jakimś nośniku, czyli kod, który może uchodzić za informację potencjalną (wymagającą zarejestrowania różnicę stanów fizycznych). Dopiero dzięki *zarejestrowaniu* kodu może zaistnieć informacja jakościowa, będąca czynnikiem realizacji informacji strukturalnej i reprezentacji psychicznych.

Słaby funkcjonalizm, jako teoria fizykalistyczna, również okazuje się teorią nieadekwatną dla opisu działania mózgu. Ważne jest bowiem to, aby przestrzenie sygnałów odróżniać od zbiorów sygnałów, czyli zbiorów pewnych stanów bądź procesów fizycznych.

Andrzej Chmielecki zauważa, że w fizykalizmie rozróżnienie to nie występuje, co pociąga za sobą utożsamienie informacji z sygnałem, czyli jej nośnikiem. W konsekwencji tego, stany przestrzeni sygnałów nie są uznawane za *informację o* rzeczywistości, lecz za odwzorowania świata zewnętrznego przypominające obrazy, które musiałyby być oglądane przez zamieszkującego głowę *homunkulusa*. Wynikiem takich poglądów jest także naiwny realizm w kwestii postrzegania rzeczywistości, do której tak naprawdę mamy dostęp tylko poprzez

[292] Trzeba zaznaczyć, że chodzi tu o sygnały nie w rozumieniu semiotycznym, lecz w rozumieniu teorii sygnałów.
[293] A. Chmielecki, *Wykłady z semiotyki*.

uzyskiwaną o niej informację, stanowiącą bazę dla konstytuowania *rzeczywistości wirtualnej*, swoistego „kosmosu wewnętrznego", w którym żyjemy.

Jak już zostało powiedziane, zaistnienie informacji wymaga zarejestrowania kodu. Dlatego też pojawienie się informacji, pojętej jako podprzestrzeń przestrzeni sygnałów, wymaga zarejestrowania danego zbioru stanów przestrzeni sygnałów przez jakąś inną populację neuronów. Odbywa się to w drodze odwzorowania jednej przestrzeni sygnałów w inną z zachowaniem pewnych istotnych relacji. Fakt pomijania w odwzorowaniu relacji nieistotnych wskazuje na to, że nie mamy do czynienia z *izomorfizmem* (wierną kopią zbioru stanów przestrzeni sygnałów), lecz *homomorfizmem*, czyli *kompresją* zbioru stanów przestrzeni sygnałów w zbiór zawierający mniej elementów, ale zachowujący istotne między nimi relacje. Taki sposób rejestrowania zbiorów stanów przestrzeni sygnałów jest uzasadnieniem faktu, że zaistnienie informacji może nastąpić nawet wtedy, gdy niektóre neurony są nieaktywne (na przykład wskutek ich uszkodzenia). Dzięki odwzorowaniom homomorficznym, milczenie neuronów staje się informacją, że nic je nie pobudza.

Według Chmieleckiego, przedstawiony wyżej sposób rejestrowania i generowania informacji przez układ nerwowy wystarcza do tego, aby informacja ta była informacją *o czymś*. Może to być informacja jakościowa lub strukturalna. Integralnie traktowany przez centralny układ nerwowy zbiór zarejestrowanych w powyższy sposób informacji zmysłowych Chmielecki nazywa *reprezentacją psychiczną*.

Interpretacja reprezentacji psychicznych u zwierząt polega na przyporządkowaniu tym reprezentacjom jakichś reakcji behawioralnych, co odbywa się w trybie odruchów bezwarunkowych (wrodzonych) lub warunkowych (wyuczonych). Odruchy warunkowe nabywane są na zasadzie asocjacji, co wiąże się z koniecznością posiadania krótkoterminowej pamięci *operacyjnej* (potrzebnej do wytworzenia asocjacji) i długoterminowej pamięci *trwałej* (pozwalającej na zapamiętanie asocjacji i rozpoznanie jej elementów)[294].

Organizm posiadający centralny układ nerwowy nie reaguje więc bezpośrednio na bodźce zewnętrzne, lecz na reprezentację psychiczną tego bodźca[295]. Organizm taki należy zaliczyć do dziedziny bytów psychicznych właśnie dlatego, że występujące w jego układzie nerwowym reprezentacje świata zewnętrznego są wykorzystywane do sterowania funkcjonowaniem tego organizmu (jest to jednoznaczne ze stwierdzeniem, że organizm taki posiada *psychikę*).

Reprezentacje świata zewnętrznego mogą powstawać na przykład w wyniku pojawienia się informacji wzrokowej, czyli zarejestrowania przez oko różnic stanów fizycznych, których nośnikiem jest światło odbite od realnego przedmiotu. Zwierzęta wykorzystują tak powstałe reprezentacje wyłącznie na podstawie zapamiętanego bądź wrodzonego związku następstwa czasowego między reprezentacjami lub ich korelacji. Informacje dostępne zwierzętom są tylko potencjalnie informacjami *o czymś*, czego zwierzęta świadome być nie mogą, ponieważ w dziedzinie bytów psychicznych nie może być mowy o świadomości. Explicite z informacją *o czymś* mamy do czynienia dopiero na poziomie bytów duchowych.

[294] Asocjacjonizm Hume'a jest w tym ujęciu teorią, która nie wykracza poza dziedzinę bytów psychicznych, czyli zwierząt.
[295] To, co dzieje się pomiędzy bodźcem a reakcją, *jest* ważne. Metaforycznie można by powiedzieć, że trzeba pokonać mrok „czarnej skrzynki".

4.2. Spostrzeżenie jako semioza i rozumienie

Człowiek posiada zdolność *spostrzegania*, czyli takiego interpretowania informacji zmysłowej, które polega na przyporządkowaniu danej reprezentacji psychicznej jakiegoś obiektu w świecie. Inaczej mówiąc: treści, które pojawiają się w układzie wzrokowym, są interpretowane jako spostrzeżenie rzeczy istniejącej obiektywnie. Dzieje się to w wyniku operacji polegających na analizie i interpretacji tego, czym układ spostrzegający już dysponuje. Interpretacja ta pozwala przyporządkować reprezentacji psychicznej jakiś przedmiot, czyli przetworzyć ją w spostrzeżenie *czegoś*. Pierwotna reprezentacja psychiczna zostaje więc potraktowana jako *znak*, co świadczy o tym, że mamy tu do czynienia z semiozą (sytuacja znakową). Mechanizm ten, oparty na kompetencji znakowej użytkownika, polega na uznaniu reprezentacji psychicznej za coś niesamodzielnego, wymagającego umocowania bytowego w postaci jakiegoś przedmiotu, który jest sensem tej reprezentacji. Działanie tego mechanizmu jest zależne od wstępnego założenia, że *coś*, co ta reprezentacja reprezentuje (*coś*, od czego zależy, że reprezentacja jest reprezentacją tego czegoś) realnie istnieje.

Przekonanie o realnym istnieniu obiektywnej rzeczywistości oddziałującej na zmysły Andrzej Chmielecki określa mianem *intuicji realności*. Pojawia się ona wskutek aktywności danego organizmu w stawiającym opór realnym świecie. Założenie realnego istnienia świata zewnętrznego jest podstawą *rozumienia*, czyli uchwycenia domniemanego umocowania bytowego reprezentacji (potraktowanej jako znak do czegoś odsyłający).

Spostrzeżenie nie ma więc charakteru wrażenia zmysłowego, lecz jest jego przedmiotową interpretacją. Nie jest ono widzeniem, lecz „wiedzeniem", czyli rozumieniem polegającym na uchwyceniu sensu zmysłowych reprezentacji psychicznych. Chmielecki zwraca uwagę na to, że fenomen spostrzegania pojawił się w przedjęzykowym stadium rozwoju gatunku ludzkiego. Wynika z tego, że procesy umysłowe miały charakter semiotyczny jeszcze zanim człowiek zaczął używać języka, jako narzędzia służącego do komunikowania się.

Przedmiot spostrzegany z założenia istnieje realnie (*jest*), a jeśli istnieje, to musi też być *jakiś*. Dlatego spostrzeżenie jakiegoś przedmiotu musi wiązać się z przypisaniem mu pewnych cech, które mają swe źródło w *treści* spostrzeżenia. Pojawienie się tej treści jest możliwe dlatego, że wizualne reprezentacje psychiczne nie tylko coś reprezentują, ale także zawierają informację prezentującą to, co reprezentowane. Należy je więc uznać za znaki, których treść (*intensja*) prezentuje, a właściwie *wyznacza* przedmiot zwany *intensjonalnym*.

Intuicja realności skłania do traktowania przedmiotu intensjonalnego, jako obiektywnie istniejącego w realnym świecie. Jest to jednak przedmiot o odmiennej naturze, niż realnie istniejąca rzecz, która jest źródłem spostrzeżenia. Rzecz realna ufundowana jest w materii fizycznej i stanowi umocowanie bytowe przedmiotu spostrzeżenia, którego fundamentem (warunkiem zaistnienia) są zmysłowe reprezentacje psychiczne.

Przedmiot intensjonalny spostrzeżenia jest konstytuowanym (wyznaczanym) przez treść spostrzeżenia *sensem*, który nie ma zdolności oddziaływania, choć może determinować. Należy go wobec tego uważać za byt *irrealny*, który nie jest bytem obiektywnym i nie istnieje w realnym świecie zewnętrznym. Nie jest on jednak także bytem czysto subiektywnym, bo jako przedmiot spostrzeżenia, jest zlokalizowany w pewnej przestrzeni; nie istnieje zatem w umyśle podmiotu, lecz jest wobec niego transcendentny.

Przedmiot spostrzeżenia jest czymś pośrednim między tym, co obiektywne, a tym, co subiektywne. Taki wirtualny byt (istniejący pomiędzy rzeczywistością realną a fikcyjną) można za Immanuelem Kantem nazwać *fenomenem*, natomiast za Edmundem Husserlem należy przypisać mu cechę *transsubiektywności*.

Jako że przedmioty spostrzeżeń są obiektami wirtualnymi, to i cała rzeczywistość, którą te obiekty tworzą, jest wirtualna. Rzeczywistość ta jest światem istot duchowych, ponieważ tylko one mają zdolność uchwytywania lub ustanawiania sensu poprzez domniemanie umocowania bytowego własnych stanów wewnętrznych, uznanych za znaki. Dlatego właśnie sens jest zasadą determinacyjną w dziedzinie bytów duchowych, zaś jego poznawanie (rozumienie) jest sposobem bycia człowieka jako osoby. Dzięki zdolności rozumienia (dokonywania semioz), ludzie mogą posługiwać się kategorią znaku i tworzyć język.

4.3. Nazwy i język

Zatem tylko istoty duchowe posiadają zdolność spostrzegania, która z kolei jest podstawowym warunkiem tego, aby móc nadawać przedmiotom spostrzeżeń *nazwy*. Posługiwanie się nazwami Andrzej Chmielecki uznaje za elementarną semiozę językową.

Spostrzeganie to duchowy akt określania sensu pewnych reprezentacji psychicznych, natomiast *nazywanie* polega na przyporządkowaniu znaków dźwiękowych lub pisanych określonym sensom spostrzeżeń bądź wyobrażeń.

Nie mamy tu do czynienia z odwzorowaniem reprezentacji psychicznych, lecz z duchowym aktem przyporządkowania reprezentacjom psychicznym innych reprezentacji[296]. Przyporządkowanie to nie odbywa się na mocy następstwa czasowego, lecz jest ustanowieniem relacji między reprezentacjami na zasadzie ich *nadbudowania*. Nazwa jest w tym ujęciu reprezentacją reprezentacji, ich superpozycją. W podobny sposób tworzone są *predykaty*, czyli wyrażenia przypisujące przedmiotom intensjonalnym jakieś cechy (intensję, treść)[297].

Tym, co zostało ustanowione, można manipulować. Zestawianie nazw i predykatów w ciągi znakowe pozwala uzyskać *znaczenia słowne* i *deskrypcje*, dzięki którym można językowo charakteryzować znaczenia (treści) innych nazw[298]. W wyniku wielokrotnego zastępowania znaków przez ciągi innych znaków powstaje relacyjna struktura w postaci *grafu*. Wierzchołkami tego grafu są przedmioty reprezentacji psychicznych (ich sensy), zaś on sam, stanowiąc pewną treść, jest znaczeniem, które wyznacza sens znaku, czyli jakiś przedmiot intensjonalny.

Znaczenie jest więc relacją międzyznakową, opartą na przyporządkowaniu jakiemuś znakowi innych znaków, natomiast wyznaczony przez treść znaku przedmiot intensjonalny należy do wirtualnej rzeczywistości pozaznakowej (irrealnej bądź idealnej). Treść znaku wyznacza

[296] Usłyszane ciągi dźwięków są tylko psychicznymi reprezentacjami fal akustycznych, czyli pewnych zdarzeń fizykalnych. Ciągi te stają się reprezentacjami reprezentacji (nazwami), gdy zostaną przyporządkowane spostrzeżeniom bądź wyobrażeniom *czegoś*, czyli innym reprezentacjom psychicznym.

[297] „Odmiennie niż w sferze animy, gdzie mamy do czynienia z jednowymiarowymi, liniowymi związkami następstwa czasowego reprezentacji i stanów wyrazowych (procesami) – w dziedzinie duchowej pojawiają się «drzewiaste», dwuwymiarowe, bezczasowe relacje znaczeniowe" – A. Chmielecki, *op. cit.*, 1999, s. 136.

[298] Na przykład *państwo związkowe* (nazwa i predykat) jest znaczeniem nazwy *federacja*.

określone cechy przedmiotu intensjonalnego, ale nie wszystkie. Graf znaczenia składa się tylko ze znaków dostępnych danemu podmiotowi i tylko te znaki mogą determinować cechy konstytuowanego przez ich strukturę przedmiotu intensjonalnego. Wyraźnie tu widać, jak wielką rolę w procesie poznania i twórczości odgrywa kompetencja znakowa użytkowników języka.

Kompetencja znakowa jest koniecznym, choć niewystarczającym warunkiem tego, że reprezentacja do czegoś odsyła i staje się przez to znakiem. To, na mocy czego odsyłanie zachodzi, Chmielecki nazywa *zasadą* znaku. Znaki mogą odsyłać do jakiegoś przedmiotu na mocy związku naturalnego (są to *oznaki*), podobieństwa (*ikony*) lub konwencji (*symbole*). Kompetencja znakowa polega na znajomości tej zasady odsyłania. Owo odsyłanie odbywa się za pośrednictwem procesu interpretacji, czyli rozumienia. Dzięki rozumieniu znak nabywa swe znaczenie i odniesienie przedmiotowe. Sam znak natomiast to całość dwuelementowa, składająca się z elementu znaczącego, czyli nośnika funkcji znakowej (*signifiant*) oraz wyżej opisanej zasady znaku[299].

Bez kompetentnego użytkownika znak zaistnieć nie może. Sytuacja znakowa obejmuje zatem – po pierwsze – *znak,* a wraz z nim jego użytkownika, czyli osobę, w której umyśle powstał stan będący reprezentacją psychiczną.

Drugim elementem sytuacji znakowej jest intensja (treść) znaku, przypisana mu poprzez sposób jego rozumienia, czyli interpretację polegającą na zastąpieniu jednych znaków przez inne znaki *nieumowne*, będące składnikami kompetencji znakowej użytkownika (jest to operacja syntaktyczna).

Znakami nieumownymi (bezpośrednio prezentującymi swój przedmiot) są ikony i niektóre oznaki. Za takie znaki można uznać reprezentacje psychiczne (zbiory informacji zmysłowych, powstające w drodze odwzorowania zachowującego pewne relacje lub na zasadzie przyporządkowania przyczynowo-skutkowego). Pojęcia i sądy przedstawiają stany rzeczy właśnie za pomocą znaków nie tylko reprezentujących, ale też prezentujących te stany.
Jeśli z kolei mówimy o znakach umownych, to poruszamy się w sferze językowej. Znaki te mogą tylko reprezentować swój przedmiot, czyli prezentować go pośrednio, poprzez przypisaną im intensję, odwołującą się ostatecznie do prezentujących bezpośrednio spostrzeżeń i wyobrażeń albo utworzonych na ich bazie pojęć i sądów. Odwołanie to jest wyjściem poza język, uchwyceniem przedmiotowego stanu rzeczy, co jest konieczne dla zrozumienia języka.

Zarówno element trzeci sytuacji znakowej, którym jest wyznaczany przez treść znaku irrealny przedmiot intensjonalny (sens), jak i element czwarty, czyli desygnat znaku (to, co znak oznacza), są elementami należącymi do rzeczywistości pozaznakowej (wirtualnej w przypadku sensu i realnej w przypadku desygnatu).

Znak oznacza realnie istniejący obiekt lub stan rzeczy tylko wtedy, gdy ten obiekt lub stan rzeczy spełnia intensję (treść) tego znaku. Spełnianie to odbywa się z dokładnością do pewnego morfizmu, wobec czego intensja znaków nie zapewnia tego, że będą one adekwatnie opisywały rzeczywistość realną. O poznaniu obiektywnej rzeczywistości decyduje więc morfizm zachodzący między cechami przedmiotu intensjonalnego a cechami domniemanego desygnatu. Stopień tego morfizmu odpowiada stopniowi poznania świata realnego, a

[299] Jak zauważa Chmielecki, zarówno de Saussure, jak i Peirce, uważali za znak coś, co w rzeczywistości jest *sytuacja znakową*, której znak jest tylko jednym z elementów konstytutywnych.

kategoria prawdziwości wiąże się ze stopniem adekwatności rzeczywistości wirtualnej do realnej.

Ciągi dźwięków i napisów, czyli utworzone poprzez manipulowanie znakami nośniki reprezentacji językowych, można łatwo wytwarzać i w drodze komunikacji udostępniać innym. Dzięki abstrahowaniu od doświadczenia, język daje możliwość poznania teoretycznego, czyli konstruowania nowych przedmiotów intensjonalnych jako pojęciowych modeli rzeczywistości.

4.4. Myślenie

Poznaniem bazowym na poziomie duchowym jest intuicja realności, którą Andrzej Chmielecki nazywa rozumieniem rzędu pierwszego. Zdolność rozumienia wyższego rzędu wymaga osiągnięcia rozumienia niższych rzędów. Rozumienie rzędu trzeciego polega na uchwyceniu jakiegoś abstrakcyjnego układu relacji, a nie konkretnego zjawiska, jak to ma miejsce w przypadku spostrzeżenia, czyli rozumienia rzędu drugiego. Dlatego też rozumienie rzędu trzeciego (uchwytywanie sensu) można uznać za nadbudowane na rozumieniu „płytszym".

W przypadku rozumienia zawsze mamy do czynienia z operacjami na reprezentacjach psychicznych potraktowanych jako znaki. Interpretacja nieznanych znaków, czyli przyporządkowanie im znaków wchodzących w skład kompetencji znakowej użytkownika, jest operacją wykonywaną przez *umysł*, który wykrywa także związki pomiędzy już zinterpretowanymi znakami.

Umysł związany jest z mechanizmami pamięci trwałej, a występujące na jego poziomie semiozy mają charakter syntaktyczny. Struktura pamięci trwałej składa się ze śladów pamięciowych reprezentacji. Operacje umysłowe „polegają na wyróżnianiu pewnego podgrafu z obszerniejszego grafu, jaki tworzy struktura pamięci trwałej (…) a następnie na przesłaniu tego podgrafu do pamięci operacyjnej"[300].

Umysł nie ma wglądu w treść reprezentacji, którymi operuje (są one dla niego czarnymi skrzynkami). Według Chmieleckiego, dostęp do treści reprezentacji psychicznych jest możliwy dopiero za pomocą *intelektu*, który operuje zasobami pamięci operacyjnej, w której mogą być konstruowane większe całości. Rozumienie rzędu trzeciego, czyli uchwytywanie sensu i przechodzenie od reprezentacji do tego, co ona oznacza, także jest rozumieniem intelektualnym. Na poziomie intelektu następuje przejście od syntaktyki (rozumienia płytszego) do semantyki (rozumienia głębszego). Dopiero tutaj mamy do czynienia z myśleniem.

4.5. Czy maszyny mogą myśleć?

Zespół struktur pełniących w organizmie funkcje kontrolowania i sterowania zachowaniem to w terminologii Andrzeja Chmieleckiego *animalny układ sterujący*. Funkcje te koordynuje *centrum sterujące* tego układu, będące jego podukładem. Animalny układ sterujący posiadają nie tylko zwierzęta, ale także ludzie. Układ ten spełnia swoje funkcje już od chwili narodzin

[300] A. Chmielecki, *op.cit.*, kognitywistyka.net.

człowieka i stanowi bazę dla wykształcenia się *duchowego* centrum sterującego. Duchowe centrum sterujące wykorzystuje zasoby informacyjne pamięci operacyjnej, które docierają do niej z poziomu animalnego, dzięki czemu centrum duchowe ma wgląd w to, co dzieje się w psychice, kontrolując i ukierunkowując jednocześnie proces myślenia.

Podstawą znaczeń i sensów są reprezentacje, które odwzorowują świat realny na zasadzie morfizmu. W animalnym układzie sterującym ma miejsce behawioralna (popędowa) interpretacja tych reprezentacji, dokonywana poprzez ich przekład na aktualne stany organizmu. Natomiast duchowy układ sterujący interpretuje reprezentacje w kategoriach przedmiotowych, a następnie podmiotowych, ustalając, co dana reprezentacja dla podmiotu oznacza i czy jest *ważna*. Interpretacja ta odbywa się poprzez odniesienie do obiektywnej rzeczywistości.

Myślenie nie jest manipulacją symbolami, odbywającą się zgodnie z jakimiś regułami określającymi syntaktyczne związki między znakami. Proces myślenia polega na rozumieniu związków znaczeniowych i związków sensu, co wymaga wyjścia ku rzeczywistości pozaznakowej (wirtualnej, wyznaczonej przez treść znaku, oraz realnej, w której istnieją rzeczy oznaczane przez znak).

Zastąpienie jednego wyrażenia ciągiem innych wyrażeń jest operacją syntaktyczną, którą mógłby wykonywać pracownik siedzący w „chińskim pokoju" Searle'a. Jednak ani ten pracownik, ani cały pokój nie mógłby wykroczyć poza formalne działania językowe i dlatego nie mógłby uchwycić sensu chińskich symboli, którymi operował. „Chiński pokój" nie tylko nie jest w stanie zrozumieć chińskich symboli. „Chiński pokój" w ogóle nie jest zdolny do rozumienia.

Jeśli uznamy, że komputer jest urządzeniem operującym symbolami, to nie wystarcza to do tego, aby mógł myśleć. W myśleniu nie chodzi bowiem o zdolność manipulowania symbolami, lecz o zdolność uchwytywania tego, *co* one symbolizują. Aby pojawił się fenomen myślenia, reprezentacje psychiczne muszą być uznane za znaki *coś* oznaczające.

Operowanie symbolami ma charakter czysto funkcjonalny. Działanie komputera również polega na spełnianiu określonych programem funkcji. Człowiek jest istotą myślącą, która operuje symbolami. Pochopne wyprowadzenie z tych faktów wniosku, że myślenie ma charakter czysto funkcjonalny, doprowadziło do sformułowania hipotez, które stały się dogmatami ideologii Sztucznej Inteligencji. Hipotezy te są fałszywe.

Po pierwsze, w drodze dokonywania formalnych operacji syntaktycznych na znakach umownych (symbolach) nie uzyska się semantyki, która wymaga przekładu tych znaków na znaki nieumowne (prezentujące, czyli zawierające informację o tym, co te znaki reprezentują).

Po drugie, myślenie pojawia się dopiero na poziomie rozumienia, czyli w sferze *pragmatyki*, a nie syntaktyki, czy semantyki. Myślenie jest tam, gdzie mamy do czynienia z podmiotem dokonującym operacji intelektualnych i przedmiotem wyznaczonym przez treść reprezentacji psychicznych. Zanim jednak poziom ten zostanie osiągnięty, reprezentacje muszą stać się znakami, gdyż same w sobie nie posiadają żadnego znaczenia. Zyskują je za sprawą działania duchowego układu sterującego.

Reprezentacje (jako znaki) są umysłowo interpretowane za pomocą pamięci operacyjnej na podstawie danych zawartych w pamięci trwałej. Komputer również posiada pamięć trwałą i operacyjną, ale w przypadku maszyny cyfrowej obie służą tylko do przechowywania zakodowanej informacji, czyli wyników operacji oraz danych.

Duchowe centrum sterujące wykorzystuje pamięć dzięki wytworzonym związkom znaczeniowym, natomiast system operacyjny komputera odnajduje komórki pamięci dzięki przypisanym im adresom. Engramy, czyli ślady pamięciowe reprezentacji, tworzą zbiory *kolektywne* (struktury te powstają wskutek zapamiętania asocjacji reprezentacji), co umożliwia nie tylko interpretację nowo pozyskanej informacji, ale także pozwala na ciągłe uczenie się i tworzenie w umyśle czegoś zupełnie nowego. Nie jest to możliwe w przypadku komputera, w którym zbiór *bajtów* (komórek pamięci) jest zbiorem *dystrybutywnym*. Wykorzystanie informacji przez komputer nie wynika z jej interpretacji, lecz z tego, że konstrukcja tej maszyny pozwala na kierowane programem przetwarzanie struktur różnic stanów fizycznych.

Komputer jest urządzeniem informacyjnym i niczym więcej. Nie ma w nim relacji semiotycznych, które występują w umyśle, nie mówiąc już o rozumieniu intelektualnym. Według Andrzeja Chmieleckiego, procesy umysłowe mają charakter syntaktyczny, ale w komputerze relacje syntaktyczne nie występują.

W komputerze nie ma nawet znaków. Choć znaki pojawiają się na poziomie programu, to są one znakami tylko dla programisty, a nie dla komputera[301]. Wszystko, do czego komputer jest zdolny, to rejestrowanie i wykorzystanie różnic stanów fizycznych. Symbolizowanie tych stanów zerami i jedynkami jest działaniem na poziomie języka. Działanie to jest ludzkie, bo wymaga myślenia. Jest też działaniem praktycznym, gdyż pozwala zaprogramować komputer do wykonania określonej pracy i ułatwia zrozumieć sposób jego działania.

Błędem jest dopatrywanie się istnienia znaków tam, gdzie w rzeczywistości mamy do czynienia tylko i wyłącznie z informacją. Komputer nie posiada umysłu, gdyż nie ma zdolności dokonywania semioz, więc nie może być też użytkownikiem znaków, a co za tym idzie, nie może używać języka. Może tylko symulować jego użycie, jeśli tylko tak zostanie zaprogramowany.

Jak sama nazwa wskazuje, komputer jest maszyną liczącą. Obliczeniowa teoria umysłu zakłada, że myślenie ma charakter komputacji, które są implementowane w mózgu. Jak twierdzi Chmielecki, jest mało prawdopodobne, aby mózg naturalnie ewoluował w taki sposób, by stać się bazą dla implementacji algorytmu. Algorytm stanowi całość, w której nie może zabraknąć żadnego z tworzących ją elementów. Zanim elementy te utworzyłyby taką całość, byłyby po prostu nieprzydatne i jako takie, musiałyby zostać wyselekcjonowane w sposób przypadkowy, a nie w drodze doboru naturalnego. Dlatego nie można twierdzić, że mózg implementuje komputacje. Można natomiast uznać, że mózg dokonuje obliczeń na poziomie informacyjnym, gdzie komputacjami są morfizmy występujące przy powstawaniu

[301] Jeśli „wystukamy" na klawiaturze komputera słowo *drzewo*, to nie będzie ono dla komputera ani znakiem (złożonym) obiektywnie istniejącego drzewa, ani zbiorem symboli (liter), ani nawet ciągiem zer i jedynek, który odpowiada temu słowu w „języku maszynowym". Komputer zarejestruje tylko zmiany stanów fizycznych i wykorzysta je tak, jak pozwala na to jego konstrukcja i nakazuje program (na przykład przetworzy informacje w ten sposób, że powstanie odpowiednich różnic stanów fizycznych na ekranie monitora sprawi, że użytkownik komputera zobaczy wyświetlone słowo *drzewo*).

reprezentacji psychicznych. Umysłowe (semiotyczne) operacje na tych reprezentacjach nie mają już charakteru komputacji, choć bazują na ich rezultatach.

Odpowiedź na pytanie zadane w tytule niniejszego podrozdziału można zawrzeć w jednym zdaniu: komputer *nie może* myśleć, ponieważ nie jest *niczym więcej*, jak tylko układem informacyjnym.

Wnioski końcowe

Dawno temu człowiek kopał, odgarniając ziemię rękami. Jego zdolności intelektualne pozwoliły mu wpaść na pomysł użycia do tego celu płaskiego przedmiotu. Nazwał go łopatą. Aby usprawnić swe działania, człowiek wykorzystał zdobytą wiedzę i zbudował koparkę, którą musiał obsługiwać. Obecnie człowiek potrafi skonstruować takie urządzenie, które może kopać bez obecności człowieka, na podstawie programu, „przenoszącego" decyzje podejmowane przez programistów na przyszłe działanie maszyny. Sprawia to wrażenie, że mamy do czynienia z myślącą koparką, choć tak naprawdę jej praca jest jedynie odbiciem myślenia jej konstruktorów i programistów.

Nie widzę powodu, aby tej być może naiwnej historyjki nie odnieść do maszyn liczących, odwołując się kolejno do palców u rąk, liczydeł, kalkulatorów analogowych, maszyn cyfrowych i wreszcie „samodzielnie" podejmujących decyzje systemów eksperckich. A może człowieka można uważać za maszynę? Jak rozróżnić *jak gdyby* myślenie od prawdziwego myślenia? Mam nadzieję, że niniejsza praca stanowić będzie wkład w proces poszukiwania odpowiedzi na tego typu pytania.

Słowo *kopać*, użyte w wyżej opowiedzianej historyjce, może wiązać się nie tylko z odgarnianiem ziemi rękami, ale także z uderzaniem w coś za pomocą nóg. Nie sądzę, że słowo *myśleć* powinno być traktowane jako równie wieloznaczne. Tymczasem tak właśnie pojmują je ideolodzy Sztucznej Inteligencji. Uważam, że pewnym można być tylko tego, iż istotą myślącą jest człowiek i na tej pewności powinno opierać się rozumienie sensu słowa *myśleć*. Narzuca się tu jednak pytanie: jak dochodzi do tego, że człowiek myśli?

Wiadomo, że każda ludzka jednostka rodzi się z czaszką wypełnioną materią, zwaną *mózgiem*. Struktura tej materii pozwala na to, że może stać się ona działającym układem, którego funkcjonowanie polega na przetwarzaniu informacji, czyli – mówiąc potocznie – materia ta może stać się mózgiem *żywym*. W toku analizy tej materii można dokonać jej podziału na kilka poziomów strukturalnych. Poziom najpłytszy może podlegać opisowi w języku potocznym, jako poziom ogólny (odnoszący się do wszystkich ludzi) i strukturalnie niezmienny (na tym poziomie wskazuje się na przykład poszczególne części mózgu, jako odpowiedzialne za dany typ osobowości).

W miarę zagłębiania się w inne poziomy, zmuszeni jesteśmy używać języka naukowego (biologia, chemia, fizyka) i dostrzegamy „plastyczność" materii mózgu, która nadaje jej cechy indywidualne (każdy człowiek ma mózg typowo ludzki, a jednak inny od pozostałych ludzkich mózgów). Stwierdzenie, że materia ta jest działającym układem, którego funkcjonowanie polega na przetwarzaniu informacji, również należy do opisu na innym poziomie, niż ten, na którym działanie to określa się mianem *życia* (albo ten, na którym sposób tego działania nazywa się *rozumieniem*).

Poziomy opisu nazywam tu *płytszymi* lub *głębszymi*, aby zaznaczyć, że są one rozdzielone i do siebie nawzajem nieredukowalne, choć w pewnym sensie „przenikają się". Płytsze poziomy opisu nadbudowane są na głębszych tak, jak dające się nimi objąć, poszczególne warstwy bytowe rzeczywistości. Byty należące do poziomów głębszych można opisywać, wykorzystując byty z poziomów płytszych (i odwrotnie)[302], ale nie może się to wiązać z redukcją jednych do drugich poprzez formułowanie definicji, w których *genus proximum* albo *differentia specifica*[303] (czy też cały definiens) utworzono z pojęć poziomu, do którego nie należy definiendum[304].

Definiowanie pojęć z poziomów płytszych za pomocą pojęć poziomu głębszego może prowadzić do takich poglądów, które pozwalają uznać człowieka za „bezduszną" maszynę albo (wskutek dokonania porównań na poziomie głębokim) dostrzec w maszynach zalążki tego, co na poziomie płytkim nazywa się *myśleniem*, *rozumieniem*, czy *życiem*[305].

Definiowanie pojęć z poziomów głębszych za pomocą pojęć poziomu płytkiego może z kolei, w przypadku kwestii dotyczących myślenia, skutkować błędami *homunkulusa*[306]. Metaforyczne stosowanie pojęć jednego poziomu do wyjaśniania zagadnień dotyczących innego poziomu (w szczególności głębszego)[307], może stanowić duże ułatwienie, o ile metafora stosowana jest świadomie, jawnie i tylko tam, gdzie faktycznie jest potrzebna. Sądzę, że ideologia Sztucznej Inteligencji, jak to bywa zwykle z ideologiami, pozbawiona jest tej niezbędnej dozy rozsądku. Skoro mowa o inteligencji, powróćmy do kwestii myślenia.

Plastyczność mózgu pozwala na kształtowanie się jego struktury pod wpływem informacji napływających z zewnątrz. *Myślenie* pojawia się równocześnie z pewnymi konkretnymi „zdarzeniami", mającymi miejsce na wszystkich głębszych poziomach (od psychicznego aż po molekularny). Problem dualizmu własności wynika z tego, że „zdarzenia" te traktuje się jako od siebie odrębne, co zmusza do poszukiwania zależności między nimi.

Myślenie nie jest tylko kojarzeniem idei, nie jest też wyłącznie procesem fizycznym. Jest ono „zdarzeniem", które można opisać na wiele różnych sposobów, na poziomie fizycznym, biologicznym, psychicznym, czy wreszcie, na najbardziej odpowiednim – intelektualnym.

Pomieszanie jednego sposobu opisu z innym może być źródłem wielu problemów, z których najważniejsze to problemy rodzące się wraz z przyjęciem założeń dualizmu substancjalnego lub redukcjonizmu i dualizmu własności. Funkcjonalizm, mający stanowić rozwiązanie tych problemów, niczego nie wyjaśnia, ponieważ jest jedynie sposobem opisu relacji przyczynowych, występujących na poszczególnych poziomach. Ponadto, funkcjonalizm

[302] Opisami takimi są na przykład matematyczne modele bytów fizycznych, w których byty należące do poziomu fizycznego opisuje się z wykorzystaniem liczb, czyli bytów inteligibilnych.
[303] Rodzaj najbliższy i różnica gatunkowa.
[304] Chyba, że ma to być definicja metaforyczna (np. *procesor to część mózgu komputera*), albo po prostu nie ma innej możliwości opisania danego bytu.
[305] Do takich wniosków dochodzą zwykle zwolennicy materializmu (koncepcji ontologicznej, zaprzeczającej istnieniu substancji duchowej) i fizykalizmu (związanego bardziej z opisem atrybutów, czy własności, niż z ontologicznymi twierdzeniami rozstrzygającymi problem istnienia – fizykalizm z góry zakłada, że istnieje tylko to, co fizyczne).
[306] Mamy z tym do czynienia w przypadku myślicieli nie mogących się uwolnić od dualizmu psychofizycznego (będącego koncepcją ontologiczną) i dualizmu własności (ugruntowanego w fizykalizmie, a więc unikającego ontologicznych twierdzeń rozstrzygających problem istnienia).
[307] Jak to czyni Daniel Dennett.

abstrahuje od ontologicznych twierdzeń dotyczących materii, nie przywiązując żadnej wagi do tego, z czego mózg jest zbudowany i jak to się stało, że coś o konsystencji owsianki może myśleć. Jest to podstawowy błąd funkcjonalizmu, zajmującego się tylko formą i starającego się w ten sposób ująć również to, co dotyczy treści. Stąd pomysł intelektu pojętego jako maszyna semantyczna, napędzana przez maszynę syntaktyczną. Warto jednak pamiętać, że tak, jak forma nie wyjaśni treści, tak syntaktyka nie wyjaśni semantyki.

Gdy mamy na uwadze fenomen rozumienia, to być może, na pewnym poziomie, materia rzeczywiście staje się mniej ważna od struktury, ale, na co zwraca uwagę Dennett, nie na każdym. Spór o sztuczną inteligencję można by sprowadzić do polemiki w sprawie identyfikacji tego poziomu.

Według obliczeniowej teorii umysłu, poziomem adekwatnym dla opisu procesów intelektualnych (myślenia), jest poziom matematyczny. Matematyka jest uniwersalna, bo przedmiotem obliczeń można uczynić cokolwiek. Opis matematyczny można zastosować w celu ukazania zarówno zmienności stanów całej populacji ludności, jak i stanów mózgu jednego człowieka.

Opis rozumienia poprzez odwołanie się do obliczalnych *stanów mózgu* jest tylko jednym ze sposobów ukazania pewnego specyficznego „zdarzenia", a właściwie tego, w czym uwidacznia się ono na jednym z poziomów opisu. Opis taki jest dopuszczalny, dopóki nie zacznie się go uważać za jedyny lub najodpowiedniejszy.

Zależność wyjaśniania od „metodologicznej wygody"[308] jest oczywiście uzasadniona, ale twierdzenia naukowe są najlepszym przykładem, że jeden poziom opisu nie odzwierciedla w pełni rzeczywistości, gdyż mimo eksperymentalnej sprawdzalności jednej teorii naukowej, istnieją już inne teorie (i zapewne pojawią się nowe!), które są w jakiejś części niezgodne z pozostałymi, a jednak równie prawdziwe. Uważa się na przykład, że fizyka Einsteina nie jest jedyną słuszną teorią i nadal stosuje się twierdzenia fizyki Newtona. Jest tak dlatego, że te dwie koncepcje odnoszą się do różnych poziomów rzeczywistości. Stworzenie jednej, całościowej teorii umysłu, określonej na jednym poziomie, byłoby redukcją podobną do uznania, że teoria Newtona jest jedyną i pełną teorią wszystkiego.

Aby niektóre „zdarzenia" zrozumieć, należy poznać ich opis na kilku poziomach, które wzajemnie się uzupełniają. Dotyczy to oczywiście również tego, co nazywamy umysłem. Uważam, że pierwszym krokiem ku zrozumieniu „zdarzeń umysłowych" jest teoria superweniencji. Sądzę jednak, iż błędną przesłanką fizykalistów superwenientnych jest to, że według nich, własnościami bazowymi są, w przypadku mózgu, własności fizyczne. Owszem, można tak powiedzieć o poszczególnych, indywidualnych mózgach, ale nie o mózgu jako takim. Architektura mózgu od czegoś przecież zależy.

Powodem zmian w budowie organizmów biologicznych są nie tylko przypadkowe mutacje, ale również czynnik nieprzypadkowy, a mianowicie dobór naturalny, swoista gra organizmu ze środowiskiem, prowadzona na zasadzie „radzisz sobie – grasz dalej". A „radzisz sobie" tym lepiej, im lepszą posiadasz reprezentację rzeczywistości, co oznacza, że architekturę mózgu determinują względy kognitywne, kryjące się w sferze możności, potencji. W ten sposób to, co *możliwe*, ma wpływ na to, co realizowane.

[308] Zależność taką sugeruje Putnam.

Na pewnym etapie rozwoju ludzkiego mózgu pojawia się *intuicja realności*, czyli „najpłytszy" sposób rozumienia, będący podstawą *myślenia*. Rozumienie wchodzi w zakres pragmatyki i wiąże się z takimi semiotycznymi pojęciami, jak znaczenie i sens. Poziom syntaktyki obejmuje operacje czysto formalne, bez uwzględnienia których wyjaśnienie, na czym polega rozumienie, także nie byłoby możliwe.

Fizykalny poziom opisu, dotyczący mózgu i tego, co się w nim dzieje, rozszerza zakres wiedzy na temat „zdarzenia", które nazywamy myśleniem. Jednakże, jak już sugerowałem, każdy z poziomów pozwala tylko na częściowy opis tego „zdarzenia". Jeśli zostało ono *nazwane* na jakimś poziomie opisu, to znaczy, że właśnie na tym poziomie należy doszukiwać się jego natury, co nie upoważnia do abstrahowania od innych poziomów. Myślenie da się wprawdzie opisać poprzez wskazanie określonych czynników predeterminacji, odwołując się do funkcjonowania mózgu, ale będzie to tylko poszerzenie opisu, uzupełnienie wiedzy, a nie „wreszcie prawdziwe i jedyne wyjaśnienie".

W pracy ukazuję, że na poziomie umysłu nie mamy jeszcze do czynienia z myśleniem, które pojawia się dopiero w sferze intelektu. Sugeruję tu, że pełny opis myślenia byłby możliwy tylko poprzez wielopoziomowe ujęcie tego procesu. Nie można zatem wyczerpująco opisać umysłu, czy intelektu, wyłącznie w terminach fizykalistycznych.

Pełnego opisu nie pozwoli też uzyskać ujęcie funkcjonalistyczne. Powtórzę tu zacytowane już w tej pracy stwierdzenie Andrzeja Chmieleckiego: „prawo determinuje fakty, choć może ono być poznane tylko poprzez znajomość faktów". Istota determinuje naturę, ale tylko poznanie natury pozwoli poznać istotę. Założenie, że istotą stanów umysłowych są ich funkcje, bez pytań o naturę tych stanów, jest niczym więcej, jak tylko hipotezą, która stała się dogmatem funkcjonalizmu i Sztucznej Inteligencji.

Warto zauważyć, że nie mamy w tym przypadku do czynienia z filozofią, lecz z ideologią w rozumieniu marksistowskim, jako społecznie uwarunkowaną „świadomością fałszywą", mającą swe źródło w tym, co zwykło się określać zdaniem *American way of life*[309]. Sztuczną Inteligencję można uznać za produkt typowo amerykański nie tylko dlatego, że podbudowany jest takimi amerykańskimi koncepcjami, jak pragmatyzm, operacjonizm, behawioryzm, czy funkcjonalizm. Przede wszystkim należy tu mieć na uwadze fakt, że ideologia Sztucznej Inteligencji oparta jest na przesadnym dowartościowywaniu bezdusznego postępu technologicznego poprzez tchnięcie w jego wytwory „duszy" pod postacią umysłu.

To właśnie ta ideologia przyczyniła się do ugruntowania poglądu, że skonstruowanie „myślących maszyn" jest możliwe, co sprzyjało pojawieniu się pytań o korzyści i zagrożenia z tej możliwości płynące. Stało się to oczywiście nie tylko podstawą dla rozważań etycznych, ale także – a właściwie przede wszystkim – okazało się wspaniałą pożywką dla amerykańskiego show-businessu, ułatwiającego zdobycie funduszy na prowadzenie dalszych badań nad sztuczną inteligencją.

[309] Jak napisał Friedrich Engels, „ideologia jest procesem dokonywanym przez tzw. myśliciela wprawdzie ze świadomością, ale ze świadomością fałszywą. Właściwe siły napędowe zostają mu nie znane, inaczej przecież nie byłby to proces ideologiczny" – cytat za: A. Wood, *Fałszywa świadomość*, w: *op. cit.*, T. Hondericha (red.), t. I, s. 233.

Bibliografia

1. Arbib M.A., *Mózg i jego modele*, tłum. S. Bogusławski, PWN, Warszawa 1977.
2. Armstrong D.M., *Materialistyczna teoria umysłu*, tłum. H. Krahelska, PWN, Warszawa 1982.
3. Arystoteles, *Dzieła wszystkie*, t. III, tłum. P. Siwek, PWN, Warszawa 1992.
4. Bense M., *Świat przez pryzmat znaku*, tłum. J. Garewicz, PIW, Warszawa 1980.
5. Block N., *The Mind as the Software of the Brain*, http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/msb.html.
6. Bobryk J., *Locus umysłu*, PAN, Wrocław 1987.
7. Bremer J., *Problem umysł-ciało*, WAM, Kraków 2001.
8. Chlewiński Z. (red.), *Modele umysłu*, PWN, Warszawa 1999.
9. Chmielecki A., *Intuicja intelektualna, fenomen rozumienia*, http://kognitywistyka.net/artykuly/ach_intuicja_intelektualna.pdf.
10. Chmielecki A., *Między mózgiem i świadomością*, PAN, Warszawa 2001.
11. Chmielecki A., *Rzeczy i wartości*, PWN, Warszawa 1999.
12. Chwedeńczuk B. (red.), *Filozofia umysłu*, Aletheia, Warszawa 1995.
13. Dennett D., *Natura umysłów*, tłum. W. Turopolski, CIS, Warszawa 1997.
14. Devlin K., *Żegnaj, Kartezjuszu*, tłum. B. Stanosz, Prószyński i S-ka, Warszawa 1999.
15. Heller M., Mączka J. (red.), *Jedność nauki - jedność świata?*, Biblos, Tarnów 2003.
16. Hodges A., *Turing*, tłum. J. Nowotniak, Amber, Warszawa 1998.
17. Honderich T. (red.), *Encyklopedia filozofii*, t. I i II, Zysk i S-ka, Poznań 1998.
18. Kartezjusz, *Rozprawa o metodzie*, tłum. T. Boy-Żeleński, De Agostini, Warszawa 2002.
19. Kasperski M.J., *Sztuczna Inteligencja*, Helion, Gliwice 2003.
20. Kim J., *Umysł w świecie fizycznym*, tłum. R. Poczobut, PAN, Warszawa 2002.
21. La Mettrie de J.O., *Człowiek-maszyna*, tłum. S. Rudniański, De Agostini, Warszawa 2003.
22. Lem S., *Tajemnica chińskiego pokoju*, http://lem.arg.pl/.
23. Lyons J., *Chomsky*, tłum. B. Stanosz, Prószyński i S-ka, Warszawa 1998.
24. Mietzel G., *Wprowadzenie do psychologii*, tłum. E. Pankiewicz, GWP, Gdańsk 1998.
25. Penrose R., *Nowy umysł cesarza*, tłum. P. Amsterdamski, PWN, Warszawa 2000.
26. Platon, *Dialogi*, t. I i II, tłum. W. Witwicki, Wyd. Antyk, Kęty 1999.
27. Popkin R.H. (red.), *Historia filozofii zachodniej*, Zysk i S-ka, Poznań 2003.
28. Putnam H., *Mind, Language and Reality*, Cambridge University Press, Cambridge 1975.
29. Putnam H., *Wiele twarzy realizmu i inne eseje*, tłum. A. Grobler, PWN, Warszawa 1998.
30. Reale G., *Historia filozofii starożytnej*, tom II, RW KUL, Lublin 1996.
31. Searle J.R., *Umysł na nowo odkryty*, PIW, Warszawa 1999.
32. Searle J.R., *Umysł, mózg i nauka*, tłum. J. Bobryk, PWN, Warszawa 1995.
33. *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/.
34. *Świat nauki*, Nr 1, Warszawa, Lipiec 1991.
35. Tatarkiewicz W., *Historia filozofii*, t. II, PWN, Warszawa 1999.
36. Zimbardo P.G., *Psychologia i życie*, PWN, Warszawa 1999.
37. Żegleń U., *Filozofia umysłu*, Adam Marszałek, Toruń 2003.