

## Wstęp do uniwersalistycznej ontologii świadomości

### Introduction to the universalist ontology of consciousness

**Abstract:** In this article I raise the question about consciousness, which stems from a philosophical issue. Distinguishes here two ontologies: the first resulting from the existence of the physical world, including biological, and the other, resulting from the assessment of the existence of non-biological beings, focusing primarily on the artificial intelligence. Ontological origins is crucial here to a finally modal treatment of such phenomena as intelligence or consciousness. I'm going to describe these phenomena as arising from these two ontologically different substrates, but in some way being similar. The consequence of this expanded understanding of intelligence and consciousness shows that they are the feature, and not the property of human.

**Keywords:** intelligence, artificial intelligence, consciousness, phenomenalism, supervenience

„Zbliżyliśmy się do ogólnego konsensusu, że pewnego dnia komputery lub roboty mogą być świadome. Innymi słowy, że nie znamy żadnego fundamentalnego prawa lub zasady w tym wszechświecie, które zakazuje istnienia subiektywnych uczuć u artefaktów zaprojektowanych lub wyewoluowanych przez ludzi”.

Za: Aleksander, Morton, 2008, s. 78

„Jeśli zaakceptuje się powszechność obliczeń, a następnie obliczenie w najszerszym tego słowa znaczeniu, można oczekiwać, że są to warunki niezbędne dla świadomości”.

Sun, Franklin, 2007, s. 152

## Inteligencja jako właściwość, a nie własność człowieka<sup>1</sup>

Przedmiotem niniejszej analizy są dwa rodzaje bytów: biologiczne oraz niebiologiczne. Na początku skupię się na zjawisku inteligencji, a następnie, stosując analogię, przejdę do zjawiska świadomości<sup>2</sup> (pojęciem świadomości posługuję się intuicyjnie, nie odwołuję się do żadnej definicji świadomości, choć najbliższe byłoby pojmowanie jej jako fenomenalnej [Franklin, Baars, Ramamurthy, 2009]). Dodatkową częścią analizy jest wskazanie na intuicyjnie narzucającą się hierarchię bytów biologicznych, dotyczącą ich inteligencji lub świadomości. W przypadku tych bytów biorę pod uwagę odróżnienie człowieka i pozostałych bytów biologicznych. Uwzględniając ten podział, dochodzę do ogólnego stanowiska dotyczącego tego, czym mogłaby być lub jest świadomość, które nazywam umiarkowanie mechanistycznym, wraz z bimodalnym podziałem na fizyczność i stany fenomenalne. Wyraża się to w przyjęciu tezy, że najważniejsze jest wewnętrzne przetwarzanie informacji<sup>3</sup> we wszelkich bytach, co jest

<sup>1</sup> Poprzez rozróżnienie właściwości i własności jakiejś cechy przynależnej jakiemuś bytowi rozumiem w przypadku właściwości, że jest to możliwość przynależenia pewnej cechy do różnych ontologicznie bytów, a w przypadku własności, że przynależność jakiejś cechy do danego bytu jest wyjątkowa, będąc determinowaną jego ontologią. Rodzajowość bytów w niniejszym artykule wyznaczona jest ich odmiennym podłożem ontologicznym: biologicznym i niebiologicznym. Gdy fenomeny takie jak inteligencja lub świadomość są traktowane jako właściwość, mogą być opisywane jako przynależne zarówno do bytów biologicznych, jak i niebiologicznych, ale gdy są traktowane jako własność bytów biologicznych, głównie mając na uwadze człowieka, wtedy są związane w sposób wyjątkowy z jego biologiczną naturą, nie przynależąc do bytów niebiologicznych.

<sup>2</sup> W niniejszym artykule rozróżnienie pomiędzy inteligencją a świadomością jest intuicyjne, sprowadzając się do akceptacji istnienia tych dwóch fenomenów, o których można stwierdzić, że różnią się od siebie jakościowo – świadomość jest traktowana jako „bycie przy zmysłach”, a inteligencja to „sprawność umysłu”. W przypadku sztucznej inteligencji można mieć na uwadze zwykle inteligentne systemy, np. programy ekspertowe lub komputery szachowe, co najczęściej nie wiąże się z ich świadomością. Takie podejście może być jednak redukcjonistyczne, negując istnienie w powyższych przykładach jakiegos rodzaju świadomości. W przypadku gdy mówimy o istnieniu świadomości w kontekście sztucznej inteligencji, wtedy można wskazywać na bardziej złożone/wyrafinowane procesy: odczuwania emocji (Ray Kurzweil), moralności (Eliezer Yudkowsky) lub kreatywności (Harold Cohen).

<sup>3</sup> Posługując się pojęciem informacji i jej przetwarzaniem, mam na uwadze tradycję płynącą z teorii systemów (cybernetyki), jaka została zaproponowana np. przez Norberta Wienera [Wiener, 1971], Marvinina Minsky'ego [Minsky, 1986] lub Mariana Mazura, który traktuje cybernetykę jako naukę „o wszelkim dzianiu się, a więc o całej rzeczywistości” [Mazur, 1976, s. 11]. Geneza dla rozumienia informacji tkwi tu także w jakościowym jej pojmowaniu [Mazur, 1970]. Odwołuję się również do związku między przetwarzaniem informacji i zjawiskiem świadomości, na jakie wskazuje David Chalmers. Z tej perspektywy proces przetwarzania informacji, jak i sama informacja są traktowane jako powszechne (należące na natury wszelkich bytów) oraz jako takie, dzięki którym powstają w procesie superwencji kolejne własności bytów: stany fenomenalne. W podobny sposób pisze Robert Stalnaker: „Natomiast przyjmuje się, że relacja superwencji jest relacją metafizyczną, a nie semantyczną: jest relacją między zbiorami własności lub faktów, a nie między teoriami czy językami. (...) teza o superwencji głosi tylko tyle, że własności z jednego zbioru [chodzi o zbiór własności fizycznych – przyp. S.M.] w pewien sposób determinują własności z drugiego zbioru [chodzi o stany fenomenalne – przyp. S.M.]” [Stalnaker, 2008, s. 101]. Podobnie ujmuję to Donald Davidson: „Definicja

podłożem dla procesów świadomych lub, jak je nazywam w części przypadków, protoświadomych. Dlatego stwierdzam, że świadomość ma podstawę w przetwarzaniu informacji, traktowanym jako proces, nad którym nadbudowują się kolejne procesy lub struktury jakościowe, co następnie wpływa na kognitywne cechy bytów. Ostatecznie stoję na stanowisku monizmu informacyjnego, który traktuję jako podstawę dla procesów świadomych, występujących w mniejszym lub większym stopniu w różnych bytach oraz tworzących świadomość w odmiennych ontologicznie podłożach: biologicznym i niebiologicznym.

Zachowując rozróżnienie między podłożem biologicznym i niebiologicznym, np. elektronicznym, jest oczywiste, że inteligencja pojawia się niezależnie od odmienności ontologicznej<sup>4</sup>, dlatego nie chodzi o powiązanie inteligencji z jakimś szczególnym sposobem istnienia, jeśli mowa o jej ontologicznym podłożu, ale o potraktowanie jej jako właściwości, która pojawia się w pewnych okolicznościach. Píše tak np. Katherine Hayles: „Z postludzkiego punktu widzenia nie ma żadnego ostatecznego rozróżnienia pomiędzy egzystencją cielesną a komputerową symulacją, biologicznym organizmem a mechanizmem cybernetycznym, ludzkimi zamiarami a teleologią robota” [Hayles, 1999, s. 3]. Podobnie rozumieją to Raymond Kurzweil i Nick Bostrom: „Z punktu widzenia moralności nie ma znaczenia, czy ktoś działa na bazie neuronów krzemowych czy biologicznych (podobnie jak nie ma znaczenia, czy masz ciemną czy też jasną skórę). Podstawy odrzucenia rasizmu czy szowinizmu gatunkowego powinny być podstawą do odrzucenia szowinizmu węglowego czy też bionizmu” [Kurzweil, 2013, s. 363, za: Bostrom, Yudkowsky, 2014]. Inteligentne zachowania bytów niebiologicznych nie muszą być izomorficzne z zachowaniem bytów biologicznych, to znaczy, że nie chodzi o traktowanie inteligencji człowieka identycznie z inteligencją komputera, lecz raczej o to, by uwzględnić, że inteligencja to nie jedynie inteligencja w postaci, w jakiej występuje u człowieka<sup>5</sup>.

---

superwencji implikuje, że zmianie własności umysłowych zawsze towarzyszy zmiana własności fizycznych” [Davidson, 2008, s. 132].

<sup>4</sup> Rozróżnienie dwóch ontologii ma charakter roboczy. Dotyczy ono analizy, w której uwzględnia się różnorodne podłoże dla zachodzenia podobnych procesów. Zamiast ‘dwie ontologie’ można było użyć pojęcia np. ‘dwóch materii’, ale uznałem, że to ostatecznie jest zbyt dookreślające dane podłoże, skupiając się na jego naturze. Używam pojęć, wskazując na odmienną ontologiczną, starając się podkreślić znaczenie procesów zachodzących w jakiejś materii, a nie jej rodzaj. Przy tym jest oczywiste, że zachowuję w domyśle podstawowe znaczenie ontologii jako ogólnej teorii bytu.

<sup>5</sup> Posługuję się pojęciem sztucznej inteligencji zbliżonym do takiego, jakie ewoluowało od jego zaproponowania w 1956 roku przez Johna McCarthy’ego [McCarthy, 1956, wersja poprawiona: 2007] (istnieją opinie, że pojęcie to pochodzi z roku 1955 i że zostało zaproponowane przez Allena Newella i Herberta Simona). Głównie chodzi o sztuczne (niebiologiczne) systemy, które zachowują się w sposób podobny do myślenia człowieka (Stan Gibilisco), takie, które są zdolne rozwiązywać problemy, jakich człowiek nie jest w stanie rozwiązać ze względu na ich złożoność za pomocą własnej inteligencji. W ujęciu Marviniego Minsky’ego sztuczna inteligencja to dziedzina dotycząca budowy maszyn, które realizują zadania przy użyciu inteligencji podobnej do inteligencji człowieka. Według Johna Haugelanda sztuczna inteligencja dotyczy tworzenia sztucznych umysłów. Ray Kurzweil określa sztuczną inteligencję jako dziedzinę badań naśladującą ludzką inteligencję w maszynie, obejmującą systemy ekspertowe, rozpoznawania obrazów, uczenia się lub rozumienia języka naturalnego.

Jedno ze źródeł tej tezy tkwi w teście Turinga [Turing, 1950], czyli w przekonaniu, że w wyniku obliczeń jest możliwe uzyskanie efektu, który będzie zachowywał wymogi semantyki [Fetzer, 2004], gdzie uznaje się, że właściwości semantyczne systemu można zredukować do syntaksy języka binarnego, jak również gdzie wymogiem staje się nieodróżnialność działania komputera od człowieka. Dodatkowo w teście tym *implicite* zawiera się, że pewne struktury powstałe na podłożu niebiologicznym fundują procesy inteligentne, odmienne od inteligencji człowieka, a w niektórych przypadkach przewyższające ludzką inteligencję. Taki byt jest bezsprzecznie inteligentny – stąd użycie pojęcia ‘sztuczna’ inteligencja nie musi być adekwatne do zjawiska inteligencji, gdyż słowo ‘sztuczna’ zawiera w sobie rozróżnienie wartościujące, a nie ontologiczne. Ontologiczne rozumienie inteligencji uwzględnia różnorodność podłoża dla inteligencji oraz to, że inteligencje można w jakiś sposób porównywać, ale ich nie wartościując na sztuczną i jakąś inną. W tym ujęciu inteligencje różnią się od siebie np. co do sposobu działania, lecz poprzez słowo ‘sztuczna’ można by rozumieć, że istnieje jakaś oryginalna lub zasadnicza, właściwa inteligencja, mając na uwadze głównie inteligencję człowieka. W przypadku rozróżniania inteligencji ze względu na jej ontologię nie chodziłoby o jej wartościowanie lub hierarchizowanie w kategoriach: sztuczna – prawdziwa, lecz o uwzględnienie tego, co można przy jej udziale zdziałać lub jak ją rozwijać. Dlatego posługuję się pojęciem inteligencji, nie rozróżniając jej na sztuczną i jakąś inną, a jedynie czasami, kiedy jest to potrzebne, wskazuję na jej ontologię – wtedy używam stwierdzenia: inteligencja na podłożu biologicznym lub niebiologicznym.

W dalszej kolejności moim celem jest przyjrzenie się z filozoficznej perspektywy procesom informacyjnym nie jedynie w aspekcie inteligencji, ale także świadomości. Chodzi o zastanowienie się nad pewnymi procesami modalnymi i podejście do zjawiska świadomości w sposób podobny, jak dzieje się to w powyżej zaprezentowanym rozumieniu inteligencji. Podobnie jak w przypadku inteligencji, nie będzie chodziło o świadomość w ludzkim sensie, lecz zaproponowanie, że zjawisko świadomości obejmuje nie tylko człowieka i że może być stosowane także do innych bytów, w tym niebiologicznych, mając na uwadze głównie złożone systemy komputerowe, czyli byty elektroniczne [Koch, Tononi, 2008].

## Ontologiczna różnorodność świadomości

Rozumienie świadomości, jakie staram się zaprezentować, ma dwa źródła. Pierwszym jest wspomniany rozwój w pojmowaniu tego, czym jest inteligencja od czasów Turinga oraz płynąca z tych badań konsekwencja z rozwijanej, np. w ostatniej dekadzie, komputacyjnej teorii umysłu [Miłkowski, 2013; Starzyk, Prasad, 2011; Sun, Franklin, 2007]. Drugim źródłem jest zaproponowanie rozumienia świadomości ze względu na wskazanie wspólnej płaszczyzny dla świadomości człowieka oraz świadomości lub protoświadomości innych, mniej złożonych bytów biologicznych, jak i w końcu bytów niebiologicznych. Świadomość, podobnie jak wspomniana powyżej inteligencja, jest tu rozumiana jako właściwość wszelkich bytów, w których zachodzą wewnętrzne procesy przetwarzania informacji [Kasperski, 2003, s. 138]. Nie chodzi

przy tym o umysł lub jego strukturę (zwłaszcza w ludzkim rozumieniu) – nie pytamy na przykład o to, czy fizyczne realizacje Maszyny Turinga tworzą jakiś rodzaj umysłu, lecz o to, czy realizacje takie mogą posiadać świadomość [Davidson, 1992, s. 203].

Można próbować podążać drogą, na której przypisuje się pewne właściwości bytom niebiologicznym, najczęściej przynależącym do człowieka. Problem leży w tym, by jednak nie narzucać tych cech, np. komputerom lub inteligentnym robotom, ale zastanowić się, czy przy pewnym stopniu ich komplikacji pewne cechy, jedynie podobne do ludzkich, nie mogłyby w nich wystąpić. Chodziłoby o pojmowanie ich działania nie tylko w takich kategoriach, które odnoszą się do urządzeń, lecz w kategoriach posiadania przez nie jakości, takich jak myślenie (lub nawet doznawanie), z tym że do ludzkich będących co najwyżej podobnymi: „Świadomość jest częścią świata przyrody. Wynika ona, jak uważamy, tylko z matematycznych i logicznych oraz z nie do końca poznanych praw fizyki, chemii i biologii; nie wynika z jakiejś magicznej lub niezemskiej jakości. To dobra wiadomość, bo to oznacza, że nie ma powodu, by świadomość nie mogła być powielona w maszynie – przynajmniej teoretycznie” [Koch, Tononi, 2008; Searle, 2007].

Istnienie tych cech jest związane z komplikacją struktury i przede wszystkim z przetwarzaniem informacji [Copeland, 2004, s. 13]. Nawet mechanistyczne podejście akceptuje podobne pojmowanie zachowań, nazywanych świadomymi – przy czym z takim, jak zaprezentowane poniżej, zredukowanym do mechanicznej fizyczności podejściem do świadomości jednak się nie utożsamiam, czyli ze stwierdzeniami, że gdy inteligentny robot odpowiednio się zachowuje, jest wtedy kognitywny, więc w tym świetle nazywany jest świadomym. Poniżej przedstawione pojmowanie świadomości bytów niebiologicznych jest tutaj punktem odniesienia dla poszerzonego rozumienia świadomości, wykraczającego poza proponowany, fizykalistyczny redukcjonizm.

„Maszyna jest świadoma [*conscious*], jeśli oprócz wymaganych mechanizmów percepcji, jej działania, uczenia się i pamięci posiada centrum wykonawcze, które kontroluje jej wszystkie procesy (świadome lub podświadome); centrum wykonawcze jest napędzane przez motywacje i wybór celów maszyny, jej przełączanie uwagi, działanie pamięci semantycznej i epizodycznej, percepcji poznawczej oraz wykorzystania rozumienia poznania i motywacji, myśli lub planów kontroli uczenia się, uwagi, motywacji i monitorowania działań. Tak więc centrum wykonawcze, odnosząc doświadczenie poznawcze wewnętrznych motywacji i planów, tworzy samoświadomość i świadomy stan umysłu” [Starzyk, Prasad, 2011, s. 263]. „(...) postaramy się zdefiniować świadomość maszyny w funkcjonalnych terminach, takich, że gdy maszyna spełnia tę definicję, to jest świadoma, bez względu na poziom lub formę świadomości, jaką może posiadać. Świadomość polega na dużo większej ilości funkcji w stosunku do inteligencji. Będzie to rozwój wyłaniających się własności maszyny i ostateczna postać świadomości będzie zależała od percepcji, pamięci, zdolności motorycznych, motywacji, myślenia, planowania itp.” [Starzyk, Prasad, 2011, s. 261].

Igor Aleksander wymienia pięć kluczowych cech bytów świadomych [Aleksander, 2009], przy czym nie ma znaczenia ontologiczne podłoże, na jakim dany byt jest ufundowany, ani jakaś szczególna jego struktura – tj. zdolność zaistnienia takiego bytu w różnych sytuacjach, zdolność wyobrażania i przypominania przeszłych zdarzeń w stosunku do aktualnego doświadczenia, posiadanie uwagi i determinacji

związanej z aktualną sytuacją, zdolność do samodzielnego wyznaczania i realizowania celów oraz zdolność do rozpoznawania emocji i dostosowywania do nich swoich zachowań. Podobne kryteria postulują Janusz Starzyk i Dilip K. Prasad, proponując definicję świadomości maszyn opartej na zdolności do zaistnienia w sytuacji rozpoznawania emocji i przewidywania następstw działań opierających się na samoorganizacji i motywacji [Starzyk, Prasad, 2011]. Jednak w niniejszym artykule staram się pokazać, że nie chodzi o takie, jak powyżej wskazane, właściwości dla rozumienia tego, czym jest świadomość, lecz o jakościowy wymiar złożoności procesów, które dokonując się – w tym przypadku – na niebiologicznym podłożu, są podstawą dla rodzaju świadomości fenomenalnej. Dlatego ten powyższy cytat, określający to, czym jest świadomość w mechanistycznym sensie, jest zbyt wąskim ujęciem, w którym nie bierze się pod uwagę, że świadomość niebiologiczna może być inna niż ludzka, co głównie wiąże się z ontologicznym podłożem. Podkreślam, że nie chodzi o zachowania kognitywne bytów elektronicznych w mechanicznym sensie, ale o hipotezę, że świadomość, podobnie jak inteligencja, ma wiele swoich postaci, w zależności od podłoża, na jakim powstaje. Porównanie mózgowi biologicznych do niebiologicznych może pokazywać ich podobieństwo [Aleksander, Morton, 2008; Mainzer, 2008], lecz z naszej perspektywy najważniejsze jest to, że informacje są tam przetwarzane w dwóch odmiennych domenach ontologicznych.

Być może w przypadku bytów niebiologicznych nie chodzi o znalezienie podobnego do biologicznego mechanizmu zawiadującego wszelkimi procesami, ale o dopuszczenie istnienia takiego procesu, który albo w ogóle, albo w części wzorowałby się na procesach biologicznych. Podobnie uważał Daniel Dennett, pisząc, że teoria świadomości musi być teorią Przedmiotu Świadomości, gdzie nie chodzi o analizę wyobrażonego centralnego systemu wykonawczego, lecz o złożoność i działanie jakiegoś stanu rzeczy, będących podstawą dla procesów świadomych [Dennett, 2001; Kurzweil, 2013, s. 260–261]. Świadomość niebiologiczna może byłaby w jakimś funkcjonalnym stopniu podobna do świadomości człowieka, ale nie o to tutaj chodzi.

Wiązanie świadomości z jakąś strukturą, a nie procesami przetwarzania informacji, może nie uwzględniać jakiegoś rodzaju świadomości. Moim celem jest pokazanie, że świadomość nie powinna być łączona wyłącznie z jakąś specjalną strukturą (jest to wtedy jakaś konkretnie ustrukturowana świadomość), ale że można ją łączyć z wszelkimi procesami informacyjnymi i różnymi strukturami. Procesy przetwarzania, które zachodzą w bardziej lub mniej złożonych strukturach, jako składowe fundowałyby świadomość, a posiadając strukturalną złożoność, wpływałyby na jakość świadomości danego bytu.

Zdeterminowanie przez jedną z ontologii, a nawet – rzecz można – ontologiczna gatunkowość, powoduje rozumienie takich właściwości jak inteligencja lub świadomość w sposób, w jaki kształtuje to dana ontologia, np. biologizm. Zasadniczą obserwacją jest to, że nie możemy dowiedzieć się, czy komputer ma stany porównywalne ze stanami świadomości człowieka, gdyż procesy w nim zachodzące, wykształcone i rozpoznawalne w formie i naturze danego bytu, na poziomie jednej ontologii, nie są wprost lub może w ogóle poznawalne dla bytów wytworzonych w ramach innej ontologii [Kurzweil, 1999, s. 370]. Nawet jeśli uznamy, że na podstawowym poziomie są to procesy takie same co do natury, to w momencie, gdy chodzi o dostrzeżone

przez człowieka odmienne ontologicznie podłoża, podobieństwo rozchodzi się i byty te przestają być rozpoznawalne na poziomie takich cech jak świadomość, która dla człowieka łączy się z biologizmem, a następnie z powstającymi na tym podłożu uczuciami lub zrozumieniem emocjonalnym [Aleksander, Morton, 2008, s. 78].

W końcu 'oni' będą nami, więc nie będzie istotnych różnic między inteligencją biologiczną i niebiologiczną. Co więcej, owe podmioty niebiologiczne będą wyjątkowo inteligentne, więc zdołają przekonać innych ludzi (biologicznych, niebiologicznych lub formy pośrednie), że są świadome. Będą one miały w zanzardzu delikatne sygnały emocjonalne, które dzisiaj przekonują nas o istnieniu ludzkiej świadomości. Będą w stanie skłonić innych ludzi do płaczu i śmiechu. Będą się również złościć, gdy inni nie zechcą przyjmować ich zapewnień [Kurzweil, 2013, s. 371].

Porównywanie świadomości biologicznej z niebiologiczną może generować błąd naturalistyczny, polegający na niemocy wyjścia poza antropocentryzm, co wynika z rozumienia świadomości w sposób, w jaki przejawia się ona u człowieka. W rezultacie utrudnia to zrozumienie, czym mogłyby być procesy świadome bytów niebiologicznych, a nawet mniej złożonych bytów biologicznych. Nie ma co oczekiwać, że świadomość komputera będzie rozpoznawalna lub łatwo rozpoznawalna na ludzki sposób. Chodzi o uwzględnienie założenia, że byty niebiologiczne mogą być świadome lub nawet odczuwać na własny, specyficzny sposób [Minsky, 2006]. Przyjęcie założenia dotyczącego istnienia świadomości w bytach niebiologicznych jedynie w kategoriach obliczeniowych powoduje, że tematyka emocji może być analizowana w czysto kognitywnych kategoriach, czyli że maszyna je rozpoznaje, ale ich nie odczuwa. Przywołana dodatkowo problematyka emocji być może mogłaby być kolejnym krokiem w analizie kognitywistycznej, która starając się odnaleźć fenomeny podobne do charakteryzujących człowieka w bytach niebiologicznych, na samym początku zakłada brak takiego doświadczenia u tych bytów. Jeśli dopuścić istnienie jakiegoś rodzaju afektów na podłożu niebiologicznym, to najpewniej nie byłyby to ludzkie lub nawet do ludzkich podobne stany, ale nie można oddalić tej tematyki na zasadzie uznania, że stany takie są przez takie byty jedynie rozpoznawalne, lecz w żadnej postaci nie mogą być ich udziałem [Yudkowsky, 2011].

Przejście od inteligencji do świadomości bytów niebiologicznych jest jakościowym i niełatwym skokiem w pojmowaniu zjawisk informacyjnych. Niełatwo jest hierarchizować świadomość, albo nawet kwantyfikować ją do samego człowieka, dopuszczając do własnej świadomości świadomość inną niż ludzka. Zapewne problem tkwi w tym, że antropocentryczne pojmowanie świadomości przeciwstawia się traktowaniu procesów informacyjnych jako świadomych, jeśli nic nie zaświadcza o nich w ludzkich kategoriach. *Jeśli komputer nie wyraża żadnych oznak świadomości, jakie są mi znane, jest oczywiste, iż jest trudno domniemywać o jego procesach świadomych, a nawet więcej, jest bardzo łatwo określać takie procesy jako czysto mechaniczne, nieposiadające innych jakości.* Dodatkowo, jeśli nawet komputer czy inteligentny robot spełniają takie kryteria wyrażania się dla człowieka poprzez różnego rodzaju aktywność, ich świadomość lub protoświadomość może być czymś całkiem innym niż zauważalne dla człowieka zachowania.

## Poszerzone rozumienie świadomości

Zastanówmy się nad zjawiskiem świadomości u bytów podobnych ontologicznie do człowieka, głównie zwierząt. Na początek przeprowadźmy doświadczenia myślowe: wiemy mniej więcej, jak interpretować zachowania bliskich człowiekowi zwierząt, takich jak psy czy koty, gdy wyrażają one emocje i dają poczucie tego, że są świadome danej sytuacji. W przypadku ptaków lub owadów nie jest to już tak oczywiste, choć przecież nie można wykluczyć, że podobne jakościowo, ale mniejsze procesy w nich zachodzą; w tym sensie niewykluczone, że zwierzęta te posiadają procesy świadome lub, jak je nazywam, protoświadome. Wyobraźmy sobie następnie granicę rozdzielającą świat nieożywiony od ożywionego. Postawmy pytanie: jakie byty miałyby być tutaj bytami granicznymi? Być może, biorąc pod uwagę głównie materię mineralną lub biologiczną, można zaproponować, że byłyby to np. krzem, żelatyna lub biomasa. Dalsze wyznaczanie granic lub hierarchizowanie bytów biologicznych rodzi coraz większe problemy, gdyż kryterium staje się typ procesów, np. synteza, metabolizm, rozrodczość [Krichmar, Edelman, 2005].

Przyjrzyjmy się sytuacji, gdy ktoś mówi o swoim kocie, że jego ukochane zwierzę rozumie to, co się do niego mówi, albo gdy stwierdza: „Mój kot odczuwa, myśli, jest świadomy danej sytuacji” i nawet „kocham mnie”. Powstaje pytanie, dlaczego są sytuacje, gdy jesteśmy skłonni zaaprobować świadomość lub uczuciowość zwierzęcia, choć nie dzieje się to na ogólnych zasadach podobnego potraktowania bytów biologicznych, lecz ich hierarchizowania bądź traktowania wybiórczo. Jeszcze trudniej jest postulować istnienie świadomości inteligentnego robota, która w powyższym sensie byłaby dalece odległa od ludzkiej natury, przez to byłaby trudna do zaakceptowania [Turkle, 2012, s. 108]. Opozycja ta szczególnie widoczna jest w sytuacji, gdy myślimy o nieświadomym działaniu komputera oraz świadomym zachowaniu np. kota (wnosząc o tym po rozpoznawalnej pod tym względem charakterystyce zachowania). Być może istnieje tu predykacja skłaniająca do uznania, że mózg elektroniczny nie jest tak zobowiązujący jak mózg działający na podłożu biologicznym. W sytuacji, gdy kot przyjdzie i wskoczy na kolana, będzie szukał kontaktu z człowiekiem, jest oczywiste, że z łatwością można stwierdzić, iż posiada jakiś rodzaj świadomości tej sytuacji oraz że posiada uczucia. Kot wyraża takie stany, o których w ludzkich kategoriach, przynajmniej częściowo, można stwierdzać. Te procesy świadomości są podobne do ludzkich, choć może nie tak złożone. Granica, poza którą należałoby przesunąć pojmowanie tego, czym jest świadomość, wydaje się w tym przypadku do przekroczenia, postulując, że kot posiada świadomość, tyle że w mniejszym stopniu niż człowiek. Czy mamy zaprzeczyć, że pewne byty biologiczne miałyby mieć świadomość, a inne, mniej złożone, nie miałyby jej mieć?

Zaczynamy od znanych przypadków ludzi, u których bardzo złożone przetwarzanie informacji prowadzi do znanych nam złożonych przeżyć. Gdy przechodzimy do systemów mniej skomplikowanych, nie wydaje się, żeby było wiele powodów, by wątpić, że psy – a nawet myszy – są świadome. Niektórzy to kwestionowali, ale myślę, że dzieje się tak często ze względu na zrównywanie świadomości fenomenalnej i samoświadomości. Myszy mogą nie mieć zbyt



wielkiego poczucia ‘ja’ i mogą nie mieć introspekcji, ale wydaje się całkiem prawdopodobne, że jest coś takiego jak ‘jak to jest być myszą’ [Chalmers, 2010, s. 482].

W takich sytuacjach, jak ta opisana powyżej, wszelkie, przynajmniej biologiczne struktury, w których zachodzą procesy przetwarzania informacji, mogłyby mieć załączki świadomości. Świadomość traktuję jako jakość przynależną do wszelkich struktur, podobnie jak traktował intencjonalność Dennett, tj. że przy założeniu, iż intencjonalność jest powszechnie charakteryzująca wszelkie byty, wśród różnych bytów można wskazać mniej lub bardziej złożone, w tym sensie mniej lub bardziej interesujące [Dennett, 1997, s. 47]. W przypadku świadomości można zastosować podobne podejście, tj. że jedne byty są bardziej, a inne mniej lub inaczej świadome. Wszystkie byty przetwarzające informacje posiadałyby świadomość, choć może w przypadku mniej złożonych bytów lepiej jest powiedzieć, jak wspomniałem powyżej, że chodzi o załączki świadomości, poczynając np. od zimującej w ziemi byliny lili, poprzez larwę motyla i samego motyla, kota, psa, do człowieka. Uważam, że świadomość jest powszechną właściwością bytów, a utożsamiana tylko ze świadomością człowieka lub nawet zwierzęcia, jest arbitralnym ograniczeniem.

(...) wszędzie tam, gdzie istnieje oddziaływanie przyczynowe, tam jest informacja, a gdzie jest informacja, tam jest przeżycie. Stany informacyjne można znaleźć w kamieniu – gdy na przykład rozszerza się i kurczy – czy nawet w różnych stanach elektronu. Jeśli więc nieograniczona zasada podwójnego aspektu jest poprawna, będą istnieć przeżycia związane z kamieniem lub elektronem [Chalmers, 2010, s. 487].

Powróćmy do niebiologicznej świadomości, tj. do pytania, czy jesteśmy skłonni przyjąć istnienie świadomości, które ma inne niż biologiczne podłoże. Ponieważ należę do grupy naukowców zgadzających się na takie potraktowanie świadomości, jestem przekonany o wspomnianym błędzie naturalistycznym, wynikającym z antropocentryzmu, polegającym na pojmowaniu świadomości jedynie w ludzkich lub biologicznych kategoriach.

Sądzę, że można poszerzyć znaczenie świadomości i przyjąć jej powszechność. Drogą do przyjęcia postulatu o uniwersalizmie świadomości jest przyjęcie takiej ontologii, która zakłada, że podłożem dla świadomości są wszelkie procesy informacyjne odnajdywane w różnych bytach. Postulat taki ma dzisiaj może niewielkie znaczenie, ale nie sądzę, by można było w przyszłości wykluczyć istnienie bytów niebiologicznych, które nie tylko będą inteligentne, ale także będą świadome i będą dawać to do zrozumienia człowiekowi [Moravec, 1999]. Zasadnicze jest tu potraktowanie w podobny sposób przetwarzania informacji, np. w takich złożonych systemach jak komputery lub ludzkie mózgi wraz z mniej złożonymi, jak np. u zwierząt. Podchodząc do świadomości jakościowo, nazywam przetwarzanie informacji podstawowym nośnikiem świadomości. Prowadzi to do postulatu poszerzonego rozumienia świadomości, tj. przede wszystkim postulowania istnienia jej w bytach innych niż człowiek. Świadomość jest w tym ujęciu stopniowalna, wyznaczona złożonością struktury, jak i sposobem przetwarzania informacji – jest powszechna i zróżnicowana. Podstawą jest

przetwarzanie informacji, ale nie ma to znaczenia dla istoty świadomości, w jakiego rodzaju materii proces ten się odbywa. Rozróżnienie ontologiczne rodzaju podłoża, nazwanego np. rodzajem materii, wpływa na rodzajowość świadomości, lecz nie na jej istnienie lub nieistnienie.

Prezentowane w niniejszym artykule stanowisko jest podobne do naturalistycznego dualizmu z prawami psychofizycznymi, regulowanymi metodologią superwencji, jakie reprezentuje David Chalmers. Poglądy te dotyczą postulatu, że świadomość można rozumieć jako powszechnie przynależną wszelkim bytom, także nieożywionym. Zwłaszcza teoria superwencji jest w tej perspektywie intuicyjnie przekonująca. Metodologia ta odpowiada na podstawowe pytanie: jak procesy mentalne powstają na gruncie procesów fizycznych? Wynika to z ontologicznego stwierdzenia, że w rezultacie zmian zachodzących w strukturze fizycznej, wiążących się z przetwarzaniem informacji, mamy do czynienia z ideami – stanami mentalnymi, odmiennymi ontologicznie od swojej fizycznej genezy: przetwarzanie informacji, jako procesów fizycznych, generuje zjawiska fenomenalne. Jak twierdzi Chalmers, ta fundamentalna teoria świadomości, zachowując spójność i komplementarność, wyjaśnia zagadnienie świadomości na poziomie ontologicznym. Jest to podejście panpsychiczne, o którym Chalmers pisze, że to rodzaj superwencji przyrodniczej, dotyczącej zjawisk należących do przyrody i wyrażających się w powszechnej relacji między tym, co fizyczne, i tym, co fenomenalne:

Możemy to wyrazić, proponując jako podstawową zasadę to, że informacja (w rzeczywistym świecie) ma dwa aspekty, fizyczny i fenomenalny. Wszędzie tam, gdzie jest stan fenomenalny, realizuje on stan informacyjny, który jest również zrealizowany w systemie poznawczym mózgu. Przynajmniej zaś dla niektórych fizycznie zrealizowanych przestrzeni informacyjnych, ilekroć stan informacyjny w tej przestrzeni jest realizowany fizycznie, jest także realizowany fenomenalnie [Chalmers, 2010, s. 470].

Przy tym Chalmers jest również obrońcą teorii mocnej SI, co powoduje, że jego poglądy można określić jako funkcjonalistyczne z uwzględnieniem dualizmu psychofizycznego wywodzącego się z teorii superwencji, co stwarza podstawy dla ogólnej teorii świadomości.

## Zakończenie

To dość komfortowa sytuacja, gdy ma się do czynienia z inteligentnym komputerem i posiada przekonanie o jego przedmiotowej naturze, nad którą góruje świadomy sytuacji człowiek. Przyjęcie faktu, że np. komputer może być świadomy, lub może lepiej: na swój sposób świadomy, może pociągać szereg konsekwencji, np. na płaszczyźnie aksjologicznej. Może to powodować, że ludzka egzystencja zostaje swoim zagrożona inną, być może nawet głębszą inteligencją lub świadomością. Problem w tym, że znamy jedynie swoją własną ludzką świadomość, przy tym nawet trudno się czasami zmierzyć ze świadomością drugiego człowieka. Sięgnięcie po naturę

świadomości w kategoriach ontologicznych pozwala na dostrzeżenie w przetwarzaniu informacji świadomości, która byłaby ciągła i dotyczyłaby całej różnorodności bytów, zmieniając się w zależności od komplikacji struktury i podłoża, na jakim występuje. Stąd wniosek, że jeśli w ogóle mamy zajmować się przetwarzaniem informacji z tej perspektywy, to nieodzowne staje się założenie ontologiczne dotyczące istnienia świadomości we wszelkich procesach informacyjnych oraz że świadomość może mieć różne oblicza. Takie cechy człowieka jak: samoświadomość, wyobraźnia, uwaga, wolicjonalność lub emocje pozostawiam poza analizą, dlatego że wyraźnie charakteryzują one człowieka i trudno jest mi w tym momencie odnieść się do nich w modalnym znaczeniu, jak również trudno jest je nazwać jedynymi wyznacznikami świadomości. Starając się zbliżyć do określenia świadomości, wychodzimy od świadomości człowieka, ale nie wiąże się to z jej rozumieniem w ramach kardynalnego sankcjonowania jakichś strukturalnych charakterystyk [Seth, 2009, s. 72]. Z tej perspektywy nie rodzajowość ontologiczna decyduje o istnieniu lub nieistnieniu świadomości, lecz rodzaj powstającej na jej gruncie struktury i przetwarzanie informacji – choć trudno pokazać, że niebiologiczne byty mają podobne do ludzkich stany fenomenalne [Kurzweil, 1999].

Znaleźliśmy się w punkcie, gdzie nie odmawia się procesów nacechowanych świadomością, w różnym stopniu, żadnym bytom. Poszerzone rozumienie świadomości zakłada modalne i różnorodne traktowanie świadomości. Powyższe ujęcie pokazuje, że załączki świadomości są we wszelkich bytach [Lloyd, Ng, 2004, s. 61].

## BIBLIOGRAFIA

- Aleksander, I. (2009). The potential impact of machine consciousness in science and engineering. *International Journal of Machine Consciousness*, 1 (1), 1–9.
- Aleksander, I., Morton, H. (2008). Computational studies of consciousness. W: R. Banerjee, B. Chakrabarti (red.). (2008). *Progress in Brain Research: Models of Brain and Mind. Physical, Computational and Psychological Approaches* (s. 77–93). Amsterdam: Elsevier.
- Bostrom, N., Yudkowsky, E. (2014). The ethics of artificial intelligence. W: K. Frankish, W.M. Ramsey, *The Cambridge Handbook of Artificial Intelligence* (s. 316–334). Cambridge University Press.
- Chalmers, D. (2010). *Świadomy umysł. W poszukiwaniu teorii fundamentalnej*. Tłum. M. Miłkowski. Warszawa: Wydawnictwo Naukowe PWN.
- Copeland, B.J. (2004). Computation. W: L. Floridi (red.). *The Blackwell Guide to the Philosophy of Computing and Information* (s. 3–17). Oxford: Blackwell Publishing Ltd.
- Davidson, D. (1992). *Eseje o prawdzie, języku i umyśle*. Tłum. B. Stanosz. Warszawa: Wydawnictwo Naukowe PWN.
- Davidson, D. (2008). Myślące przyczyny. Tłum. M. Iwanicki. W: E. Morawska (red.). *Analityczna metafizyka umysłu* (s. 132). Warszawa: Wydawnictwo IFiS PAN.
- Dennett, D.C. (1997). *Natura umysłów*. Tłum. W. Turopolski. Warszawa: Wydawnictwo CiS.
- Dennett, D.C. (2001). Are we explaining consciousness yet? *Cognition*, 79, 221–237.
- Fetzer, J.H. (2004). The Philosophy of AI and its Critique. W: L. Floridi (red.), *The Blackwell Guide to the Philosophy of Computing and Information* (s. 119–134). Oxford: Blackwell Publishing Ltd.

- Franklin, S., Baars, B.J., Ramamurthy, U. (2009). A Phenomenally Conscious Robot? *APA Newsletters – Newsletter on Philosophy and Computers*, 8 (1).
- Hayles, N.K. (1999). *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press.
- Kasperski, J. (2003). *Sztuczna Inteligencja*. Gliwice: Helion.
- Koch, Ch., Tononi, G. (2008). Can machines be conscious? *IEEE Spectrum*, The Singularity. Special Report, June, 55-56.
- Krichmar, J.L., Edelman, G.M. (2005). Brain-Based Devices for the Study of Nervous Systems and the Development of Intelligent Machines. *Artificial Life*, 11 (1), 63–77.
- Kurzweil, R. (1999). *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. New York: Viking Press.
- Kurzweil, R. (2013). *Nadchodzi Osobliwość. Kiedy człowiek przekroczy granice biologii*. Tłum. E. Chodkowska, A. Nowosielska. Warszawa: Kurhaus Publishing.
- Lloyd, S., Ng, Y.J. (2004). Black Hole Computers. *Scientific American*, 291 (5), 52–61.
- Mainzer, K. (2008). The emergence of mind and brain: an evolutionary, computational, and philosophical approach. W: R. Banerjee, B. Chakrabarti (red.). *Progress in Brain Research: Models of Brain and Mind. Physical, Computational and Psychological Approaches* (s. 115–132). Amsterdam: Elsevier.
- Mazur, M. (1970). *Jakościowa teoria informacji*. Warszawa: Wydawnictwa Naukowo-Techniczne.
- Mazur, M. (1976). *Cybernetyka i charakter*. Warszawa: Państwowy Instytut Wydawniczy.
- McCarthy, J. (1956). *What Is Artificial Intelligence?* Stanford University (wersja poprawiona: 2007).
- Miłkowski, M. (2013). *Explaining the Computational Mind*. Cambridge, MA–London: The MIT Press.
- Minsky, M. (1986). *The Society of Mind*. New York: Simon & Schuster.
- Minsky, M. (2006). *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster.
- Moravec, H. (1999). *Robot, Mere Machine to Transcendent Mind*. Oxford University Press.
- Searle, J.R. (2007). Biological naturalism. W: M. Velmans, S. Schneider (red.). *The Blackwell Companion to Consciousness* (s. 325–334). Oxford: Blackwell Publishing Ltd.
- Seth, A. (2009). The Strength of Weak Artificial Consciousness. *International Journal of Machine Consciousness*, 1 (1), 71–82.
- Stalnaker, R. (2008). Odmiany superweniencji. W: E. Morawska (red.). *Analizyczna metafizyka umysłu*. Tłum. T. Ciecierski. Warszawa: Wydawnictwo IFiS PAN.
- Starzyk, J., Prasad, D.K. (2011). Computational Model of Machine Consciousness. *International Journal of Machine Consciousness*, 3 (2), 255–281. DOI: <http://dx.doi.org/10.1142/S1793843011000819>.
- Sun, R., Franklin, S. (2007). Computational Models of Consciousness: A Taxonomy and Some Examples. W: Ph.D. Zelazo, M. Moscovitch, E. Thompson (red.). *The Cambridge Handbook of Consciousness* (s. 151–174). New York: Cambridge University Press.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59 (236) (Oct.), 433–460.
- Turkle, S. (2012). Nadejście kultury robotycznej. *Sztuka i Filozofia*, 41, 108–126.
- Wiener, N. (1971). *Cybernetyka, czyli sterowanie i komunikacja w zwierzęciu i maszynie*. Tłum. J. Mieścicki. Warszawa: Państwowe Wydawnictwo Naukowe.
- Yudkowsky, E. (2011). Artificial Intelligence as a Positive and Negative Factor in Global Risk. W: N. Bostrom, M. Čirković (red.). *Global Catastrophic Risks* (s. 308–345). New York: Oxford University Press.