

DAVID CHALMERS

MATRIKS JAKO METAFIZYKA^{*1}

1. MÓZGI W KADZIACH

Film *Matrix* przedstawia wersję starej, filozoficznej opowieści, znanej jako mózg w kadzi. Zgodnie z nią w pewnym laboratorium naukowym stoi kadź, w której pływa pozbawiony ciała mózg. Naukowiec, który nadzoruje badanie, dopilnował, by mózg był stymulowany bodźcami zewnętrznymi takiego samego rodzaju, jakie otrzymuje on normalnie, czyli wtedy, gdy jest połączony z ciałem². W tym celu mózg został podłączony do ogromnej, komputerowej symulacji świata, która określa, jakie otrzymuje on bodźce.

Prof. DAVID CHALMERS – profesor filozofii i dyrektor Center for Mind, Brain, and Consciousness na New York University oraz professor filozofii w Australian National University; adres do korespondencji: Department of Philosophy, 5 Washington Place, New York, NY 10003; e-mail: chalmers@nyu.edu

* Tytuł oryginalny: *The Matrix as Metaphysics*. Pierwotna wersja tekstu zamieszczona została na niedostępnej już stronie internetowej filmu *Matrix*. Następnie artykuł został przedrukowany w tomie *Philosophers Explore The Matrix*, red. Christopher Grau (Oxford: Oxford University Press, 2005), 132–176. Obecnie można go znaleźć pod adresem: <http://consc.net/papers/matrix.html>. Przekład za zgodą Autora. (Przyp. tłum.).

¹ Artykuł ten został opublikowany na oficjalnej stronie internetowej filmu *Matrix*, w sekcji filozoficznej. Napisałem go w taki sposób, by większa jego część była przystępna dla czytelników niezaznajomionych z filozofią. Zarazem jednak jest to poważna praca filozoficzna, nawiązująca do centralnych zagadnień z zakresu epistemologii, metafizyki, filozofii umysłu i filozofii języka. Nawiązania te wskazuję *explicite* w ostatnim punkcie, zatytułowanym „Uwagi filozoficzne”.

² Chalmers używa w tym miejscu terminologii funkcjonalistycznej, pisząc o wejściach i wyjściach (*inputs* i *outputs*). Ponieważ przy jej zastosowaniu tekst brzmiałby sztucznie, zdecydowałem się na przełożenie tych terminów jako bodźca (*input*) i reakcji (*output*), nawiązując tym samym do faktu, że funkcjonalizm wyrasta z behawioryzmu. (Przyp. tłum.).

Ponadto, jeżeli mózg wytwarza jakieś reakcje, są one uwzględniane w symulacji. Mimo że brakuje mu ciała, wewnętrzne stany mózgu zachodzą całkowicie normalnie. Z jego punktu widzenia wszystko wygląda tak samo jak dla Ciebie czy dla mnie.



Wydaje się jednak, że mózg ulega ogromnemu złudzeniu. Jego przekonania o świecie są bez wyjątku fałszywe. Sądzi on, że posiada ciało, podczas gdy tak naprawdę go nie ma. Jest przekonany, że spaceruje w promieniach słońca, lecz zamknięto go w ciemnym laboratorium. Uważa, że znajduje się w jednym miejscu, podczas gdy może być zupełnie gdzie indziej; wydaje mu się na przykład, że jest w Tucson, a tak naprawdę przebywa w Australii albo nawet w przestrzeni kosmicznej.

W takiej właśnie sytuacji znajduje się Neo na początku *Matriksa*. Jest on przekonany, że żyje w mieście, że ma włosy, że jest rok 1999 i że za oknem świeci słońce. W rzeczywistości unosi się w przestrzeni, nie ma włosów, jest rok 2199, a świat skrywają mroki wojny. Choć pomiędzy filmem a opowieścią o kadzi zachodzą drobne różnice – mózg Neo znajduje się w ciele, a komputerową symulację kontrolują nie naukowcy, lecz maszyny – najistotniejsze szczegóły pozostają te same. Krótko mówiąc: Neo to mózg w kadzi.

Przyjmijmy, że *matriks* (pisany małą literą) to sztucznie zaprojektowana, komputerowa symulacja świata³. Filmowy *Matriks* stanowi zatem jeden z przykładów *matriksa*. Będziemy również mówić, że ktoś jest zamknięty w kadzi zamiennie ze stwierdzeniem, że ktoś znajduje się

³ W tekście przyjmuję zasadę, że tam, gdzie Chalmers mówi o *matriksie* jako terminie ogólnym, zapisuję go małą literą. Jeżeli termin ten odnosi się do filmowego *matriksa*, zapisuję go wielką literą (z wyjątkiem nazw własnych, takich jak *Hipoteza Matriksa*). Natomiast tam, gdzie *Matriks* funkcjonuje jako tytuł filmu, używam kursywy. (Przyp. tłum.).

w matriksie, o ile posiada on system poznawczy, który odbiera bodźce i jakoś na nie reaguje⁴. W tym znaczeniu zamknięty w kadzi jest zarówno Neo, jak i mózg przywołany na początku tego tekstu.

Matriks mógłby symulować całą fizykę jakiegoś świata poprzez monitorowanie ruchu każdej cząsteczki w czasie i przestrzeni. (Później zbadamy również inne możliwości). Zamknięta w kadzi istota posiadałaby symulację konkretnego ciała. Połączenie z nim mogłoby być tak ustanowione, by za każdym razem, kiedy ciało otrzymuje bodziec zmysłowy, zamknięty w kadzi system poznawczy odbierał bodziec tego samego rodzaju. Natomiast kiedy system wytwarzałby ruchową reakcję, pojawiałaby się ona w układzie motorycznym symulowanego ciała.

Wraz z kwestią możliwości matriksa natychmiast pojawia się również pytanie: skąd wiem, że ja sam nie jestem w nim uwięziony? Zamknięty w kadzi, podłączony do matriksa mózg mógłby mieć przecież tę samą strukturę, co mój, przez co byłby on obdarzony przeżyciami nieodróżnialnymi od tych, których teraz doświadczam. Patrząc „od wewnątrz”, nie da się z pewnością stwierdzić, że nie jestem w takiej właśnie sytuacji. Wygląda więc na to, że nie ma sposobu, bym z całą pewnością wiedział, że nie znajduję się w matriksie.

Hipotezę, zgodnie z którą od zawsze żyję w matriksie, nazwijmy *Hipotezą Matriksa*. Głosi ona zatem, że od zawsze jestem zamknięty w kadzi. Nie jest ona jednak tożsama ze stwierdzeniem, że jestem w Matriksie, gdyż Matriks to tylko jedna z egzemplifikacji matriksa. Na razie zignorujmy komplikacje, które pojawiają się w filmie, na przykład możliwość podróży między Matriksem a zewnętrznym światem; pomijając te kwestie, możemy nieformalnie stwierdzić, że zgodnie z *Hipotezą Matriksa* jestem w tej samej sytuacji, co ludzie, którzy od zawsze znajdują się w Matriksie.

Hipoteza Matriksa domaga się poważnego potraktowania. Jak wskazał Nick Bostrom, nie jest wykluczone, że postęp technologiczny na przestrzeni dziejów wszechświata pozwoli na tworzenie komputerowych symulacji całych światów. Istnieć może ogromna liczba takich symulacji i tylko jeden świat rzeczywisty. O wiele więcej stworzeń może zatem żyć w matriksie niż

⁴ Na określenie kogoś zamkniętego w kadzi Chalmers używa terminu *envatted*. Analogicznie do innych złożań tego typu, np. *entitled* („zatytułowany”) czy *enclosed* („zamknięty”), powinniśmy mówić, że ktoś jest „zakadziowany”. Neologizm ten jest jednak niezręczny, stąd używam nieco dłuższej wersji: „zamknięty w kadzi”. Różnice między „mózgiem w kadzi” (*brain in a vat*) a kimś „zakadziowanym” czy „zamkniętym w kadzi” (*envatted*) okażą się istotne w dalszej części tekstu. (Przyp. tłum.).

poza nim. Uwzględnivszy ten fakt, moglibyśmy nawet wyprowadzić wniosek, że istnieje większe prawdopodobieństwo, że żyjemy w matriksie niż że znajdujemy się poza nim. Bez względu na to, czy mamy rację, z pewnością nie możemy być pewni, że nie jesteśmy zamknięci w matriksie.

W konsekwencji mój uwięziony w kadzi odpowiednik żyje w nieustannej iluzji. Choć wydaje mu się, że jest w Tucson, że siedzi przy biurku i pisze artykuł oraz że ma ciało, każde z tych przekonań jest fałszywe. Analogicznie, jeżeli to ja jestem zamknięty w kadzi, moje przekonania również są fałszywe: nie znajduję się bowiem w Tucson, nie siedzę przy biurku, a nawet mogę nie mieć ciała. Jeżeli zatem nie wiem, czy jestem zamknięty w kadzi, nie wiem również, czy jestem w Tucson, czy siedzę przy biurku ani czy mam ciało.

Hipoteza Matriksa zagraża praktycznie całej mojej wiedzy. Ma ona zatem charakter sceptyczny: nie mogę jej wykluczyć, a zarazem – jeżeli byłaby prawdziwa – falsyfikowałaby większość moich przekonań⁵. Tam, gdzie pojawia się hipoteza sceptyczna, żadne przekonanie nie może zyskać statusu prawdziwej wiedzy. Oczywiście, jeżeli szczęśliwym trafem nie znajduję się w kadzi, moje przekonania są prawdziwe, ale w obecnej sytuacji nie mogę wykluczyć, że są fałszywe. Hipoteza sceptyczna prowadzi więc do sceptycyzmu odnośnie do przekonań: jestem przekonany o pewnych rzeczach, ale o nich nie wiem.

Podsumowując: nie wiem, czy jestem w matriksie. Jeżeli tak, to prawdopodobnie nie znajduję się w Tucson. Skoro zatem nie wiem, czy jestem w matriksie, nie wiem również, czy jestem w Tucson. Podobnie rzecz się ma prawie ze wszystkim, co uważam za swoją wiedzę na temat zewnętrznego świata.

⁵ Przymiotnik *skeptical*, który oddaję jako „sceptyczny”, zwyczajowo przyjęło się przekładać jako „sceptyczny”. Uznałem jednak, że w tym wypadku wyłamanie się z tradycji usprawiedliwiają względy stylistyczne – wyrażenia takie jak „hipoteza sceptyczna” czy „charakter sceptyczny” brzmią osobliwie, głównie ze względu na potoczne rozumienie słowa „sceptyczny”. Ponadto, choć wyrażenia takie jak „hipoteza sceptyczna” można rozumieć jako hipotezę przyjmowaną przez sceptyków, Chalmers to właśnie ma na myśli – wszak większość rozważanych przez niego argumentów to uwspółcześnione wersje klasycznych tropów. Ostatecznie niedoskonały wydaje się zarówno termin „sceptyczny”, jak i „sceptyczny”, wybór zaś jednego z nich motywowany był wspomnianymi wyżej względami stylistycznymi. Za wskazanie tej kwestii dziękuję anonimowemu recenzentowi. (Przyp. tłum.)

2. PONOWNY NAMYSŁ NAD MÓZGAMI W KADZIACH

Jest to typowy sposób rozumienia scenariusza z kadziami. Wydają się go również przyjmować autorzy *Matriksa*. Na pudełku wydania DVD możemy przeczytać:

PERCEPCJA: nasz codzienny świat jest realny.

RZECZYWISTOŚĆ: ten świat to sztuczka, skomplikowany miraż kreowany przez potężne, kontrolujące nas maszyny.

Takie postawienie sprawy jest błędne. Uważam, że nawet jeżeli jestem uwięziony w matriksie, mój świat jest całkowicie realny. Mózg w kadzi (o ile cały czas się w niej znajduje) wcale nie ulega ogromnemu złudzeniu. Sądy Neo na temat zewnętrznego świata nie są całkowicie fałszywe. Wprost przeciwnie – przekonania istot w kadzi na temat ich świata są w znacznej mierze poprawne. Jeżeli tak, to Hipoteza Matriksa nie jest sceptyczna, a jej dopuszczalność nie podważa całej wiedzy, którą posiadam.

Podobny pogląd pojawiał się już w przeszłości. Osiemnastowieczny filozof George Berkeley utrzymywał, że rzeczywistością jest to, co mi się jawi. (Przypomnijmy sobie słowa Morfeusza: „Czym jest to, co realne? Jak to zdefiniujesz? Jeżeli mówisz o tym, czego dotykasz, co wąchasz, co smakujesz czy widzisz, to rzeczywistość sprowadza się do sygnałów elektrycznych, interpretowanych przez twój mózg”). Jeśli to prawda, świat zamkniętych w kadzi istot jest całkowicie realny: to, co im się jawi, stanowi ich rzeczywistość. W rezultacie nawet one dysponują prawdziwymi przekonaniemiami o świecie.

Niedawno odkryłem, że zgadzam się z tym wnioskiem, jakkolwiek z nieco innych powodów. Nie sądzę, by dało się obronić pogląd, że zjawiska są tym samym, co rzeczywistość, stąd trudno mi uznać rozumowanie Berkeleya. Do niedawna wydawało mi się również oczywiste, że przekonania mózgow w kadzi są w większej części fałszywe. Teraz jednak jestem zdania, że można wskazać linię argumentacji, udowadniającej, że jest inaczej.

Nadal uważam, że nie można wykluczyć hipotezy, iż żyję w matriksie. Ale nawet jeśli tak jest, nie zmienia to faktu, że znajduję się w Tucson albo że siedzę przy biurku. Hipoteza ta nie jest zatem sceptyczna. To samo dotyczy Neo – jeśli na początku filmu myśli „mam włosy” czy „na zewnątrz jest słonecznie”, nie myli się. Podobnie dzieje się, rzecz oczywista, z mózgiem w kadzi – jego myśli, np. „mam ciało” czy „chodzę”, są całkowicie poprawne.

W pierwszej chwili pogląd ten może wydać się sprzeczny z intuicją, a przynajmniej taki wydawał się mnie. Dlatego przedstawię teraz argumentację, która przekonała mnie o jego prawdziwości.

3. HIPOTEZA METAFIZYCZNA

Jak będę dowodził, hipoteza, że jestem zamknięty w kadzi, ma charakter metafizyczny, a nie sceptyczny. Odnosi się ona zatem do natury tego, co leży u podstaw rzeczywistości.

Metafizykę, w przeciwieństwie do fizyki, która zajmuje się procesami mikrofizycznymi, stanowiącymi podstawę bardziej złożonej rzeczywistości, interesuje fundamentalna natura świata. Twierdzenia metafizyczne dotyczą zatem rzeczywistości leżącej u podstaw fizyki. Mogą one również stwierdzać coś o naturze naszych umysłów albo o stworzeniu naszego świata.

Uważam, że Hipoteza Matriksa jest pewną wersją Hipotezy Metafizycznej i powinna składać się z tych trzech części. Musi ona zatem twierdzić coś o rzeczywistości leżącej głębiej niż ta fizyczna, o naturze naszych umysłów oraz o stworzeniu świata.

Hipotezę Matriksa najlepiej rozumieć zatem jako Hipotezę Metafizyczną, składającą się z następujących twierdzeń. Po pierwsze, u swych podstaw procesy fizyczne są obliczeniowe⁶. Po drugie, nasze systemy poznawcze można oddzielić od procesów fizycznych, pomimo ich wzajemnego oddziaływania. Po trzecie, rzeczywistość fizyczna została stworzona przez istoty spoza fizycznej czasoprzestrzeni.

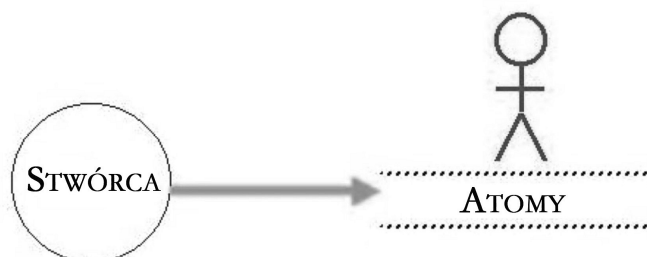
Co ważne, żadne z tych twierdzeń nie ma charakteru sceptycznego. Hipoteza Metafizyczna mówi o procesach, które leżą u podstaw naszej rzeczywistości, lecz nie wynika z niej, że rzeczywistość ta nie istnieje. Wciąż mamy ciała, nadal otaczają nas krzesła i stoły; ich fundamentalna natura jest po prostu nieco inna niż nam się zdawało. Pod tym względem przypomina ona hipotezy fizyczne, wykorzystywane na przykład w mechanice kwantowej: obie stwierdzają coś o procesach konstytuujących krzesła, choć nie wynika z nich, że krzesła nie ma. Dają nam one raczej wiedzę o tym, czym krzesła naprawdę są.

⁶ Chalmers posługuje się terminem „procesy obliczeniowe” (*computational processes*). Przymiotnik ten pochodzi od czasownika „obliczać” (*to compute*). Chalmers, prócz *computational*, używa też wyrażen *compute*, *computation* czy *computationalism*. W zależności od kontekstu, czasownik *to compute*, stanowiący źródłosłów tych wyrażen, będę tłumaczył jako „obliczać” (np. „procesy obliczeniowe”) lub jako „komputować” (np. „komputacjonizm”). (Przyp. tłum.).

Aby tego dowieść, omówię każde z trzech twierdzeń Hipotezy Metafizycznej. Jak będę wskazywał, są one spójne i nie mogą być jednoznacznie wykluczone. Ponadto nie mają one charakteru sceptycznego: większość naszych potocznych przekonań wciąż pozostaje poprawna, nawet jeśli twierdzenia te okażą się prawdziwe. To samo dotyczy ich koniunkcji, która – jak będę argumentował – jest równoważna Hipotezie Matriksa.

(1) HIPOTEZA STWORZENIA

Stwierdza ona, że czasoprzestrzeń fizyczna oraz to, co znajduje się w jej obrębie, zostały stworzone przez istoty spoza niej.

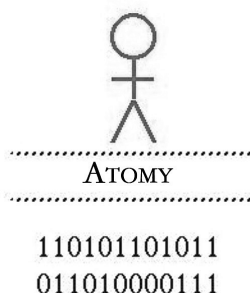


Hipoteza ta brzmi znajomo; jakąś jej wersję akceptuje wielu, może nawet większość, ludzi. Przyjmuje ją każdy, kto wierzy, że to Bóg stworzył świat oraz że znajduje się On poza fizyczną czasoprzestrzenią. Nie trzeba jednak wiary w Boga, by uznać tę hipotezę za prawdziwą. Załóżmy, że nasz świat został stworzony przez całkiem zwyczajne istoty, żyjące we wszechświecie „o piętro wyższym”, wykorzystujące najnowsze zdobycze techniki z zakresu tworzenia światów. W takim wypadku Hipoteza Stworzenia również byłaby prawdziwa. I choć nie wiem, czy jest ona prawdziwa, nie mogę również orzec, że jest z pewnością fałszywa. Jako spójna, nie może zostać jednoznacznie wykluczona.

Hipoteza Stworzenia nie jest jednak sceptyczna: nawet jeśli jest ona prawdziwa, większość moich potocznych przekonań daje się utrzymać. Wciąż mam zatem dłonie, przebywam w Tucson itd. Możliwe wprawdzie, że część z nich okaże się fałszywa, na przykład jeżeli jestem ateistą albo wierzę, że cała rzeczywistość powstała podczas Wielkiego Wybuchu. Niemniej jednak, większość moich zdroworozsądkowych przekonań dotyczących zewnętrznego świata uchodzi bez szwanku.

(2) HIPOTEZA OBLICZENIOWA

Głosi ona, że zachodzące w czasoprzestrzeni procesy mikrofizyczne są konstytuowane przez leżące u ich podstaw procesy obliczeniowe.



Wynika stąd, że znana nam fizyka nie dociera do podstawowego poziomu rzeczywistości. Tak jak u podstaw procesów biologicznych stoją procesy chemiczne, a u podstaw procesów chemicznych stoją procesy mikrofizyczne, tak coś stoi u podstaw tych ostatnich. Pod poziomem kwarków, elektronów i fotonów znajduje się głębszy poziom: bitów. Bity działają zgodnie z algorytmem obliczeniowym, który na wyższym poziomie wytwarza procesy, znane jako cząstki elementarne czy siły.

Mimo że Hipoteza Obliczeniowa nie ma tylu zwolenników, co Hipoteza Stworzenia, niektórzy traktują ją poważnie. Ed Fredkin, jej najbardziej znany sympatyk, wysunął tezę, że Wszechświat jest u swych podstaw pewnego rodzaju komputerem. Niedawno Stephen Wolfram wykorzystał ten pomysł w książce *A New Kind of Science* i zasugerował, że na najbardziej fundamentalnym poziomie rzeczywistość fizyczna może być zbiorem automatów komórkowych oraz oddziałujących na siebie bitów, działających zgodnie z prostymi regułami. Niektórych fizyków interesuje możliwość obliczeniowego formułowania praw fizycznych albo wyprowadzenia ich jako konsekwencji zasad obliczeniowych.

Można się zastanawiać, czy czyste bity istotnie mogłyby tworzyć fundamentalny poziom rzeczywistości: bit to po prostu 0 albo 1, a rzeczywistość nie może składać się z zer i jedynek. Ponadto bit to jedynie „czysta różnica” pomiędzy dwoma stanami wyjściowymi, stąd nie jest możliwe, by rzeczywistość zbudowana była z tego rodzaju czystych różnic. Bity muszą być zaimplementowane w stanach bardziej podstawowych, takich jak wolty w przypadku zwykłego komputera.

Nie wiem, czy zarzut ten jest zasadny. Trudno z całą pewnością stwier-

dzie, że wszechświat czystych bitów jest niemożliwy. To zresztą nie jest dla nas istotne. Możemy założyć, że istnieje poziom bardziej podstawowy, w którym zaimplementowane są procesy obliczeniowe. Jego natura nie ma dla nas znaczenia. Interesuje nas tylko to, że procesy mikrofizyczne są konstytuowane przez procesy obliczeniowe, które same są konstytuowane przez zjawiska bardziej fundamentalne. Od teraz, mówiąc o Hipotezie Obliczeniowej, będę miał na myśli takie właśnie twierdzenie.

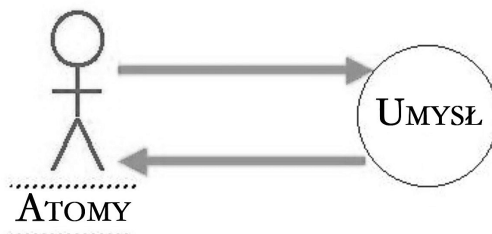
Nie jestem pewny, czy Hipoteza Obliczeniowa jest prawdziwa. Z drugiej strony, nie wiem również, czy jest fałszywa. Choć spekulatywna, jest ona spójna i nie można jej jednoznacznie wyeliminować.

Hipoteza Obliczeniowa nie ma charakteru sceptycznego. Jej prawdziwość nie powoduje, że elektrony i protony przestają istnieć, choć stają się one podobne do pierwiastków, składających się z czegoś bardziej podstawowego. Bez zmian istnieją również stoły i krzesła, a także cała rzeczywistość wyższego poziomu. Okazuje się po prostu, że poziom fundamentalny jest nieco inny niż dotąd myśleliśmy.

Przypadek Hipotezy Obliczeniowej jest analogiczny do tego, co dzieje się w fizyce kwantowej czy teorii względności, które zmuszają nas do zrewidowania niektórych przekonań „metafizycznych”, dotyczących świata: tego, że składa się on z klasycznie rozumianych cząstek elementarnych albo że czas jest własnością absolutną. Nasze zwyczajne przekonania pozostają jednak niezmiennic. Przyjęcie tej hipotezy prowadzi wprawdzie do porzucenia pewnych twierdzeń metafizycznych, na przykład przeświadczenia, że protony i elektrony to cząstki fundamentalne, ale nie wpływa na większość naszych potocznych twierdzeń.

(3) HIPOTEZA UMYŚL-CIAŁO

Zgodnie z nią mój umysł jest (i zawsze był) konstytuowany przez procesy spoza fizycznej czasoprzestrzeni, z której jednak otrzymuje on bodźce zmysłowe oraz na którą reaguje.



Hipoteza ta jest dobrze znana i zazwyczaj akceptowana. Bronił jej już Kartezjusz; jego zdaniem posiadamy niefizyczne umysły, które wchodzą w interakcje z naszymi ciałami. Wprawdzie dziś pogląd ten ma mniej zwolenników niż w czasach Kartezjusza, wciąż jednak wielu ludzi jest przekonanych o jego prawdziwości.

Bez względu na to, czy Hipoteza Umysł-Ciało jest prawdziwa, bez wątpienia jest ona spójna. Nawet jeśli współczesna nauka sugeruje czasem jej fałszywość, nie da się jej jednoznacznie wykluczyć.

Ta hipoteza również nie jest sceptyczna. Nawet jeśli mój umysł znajduje się na zewnątrz fizycznej czasoprzestrzeni, wciąż mam ciało, wciąż jestem w Tucson itd. Jej akceptacja może co najwyżej prowadzić do rewizji niektórych twierdzeń metafizycznych, dotyczących umysłu. Nasze potoczne przekonania na temat zewnętrznej rzeczywistości pozostają w większej części nienaruszone.

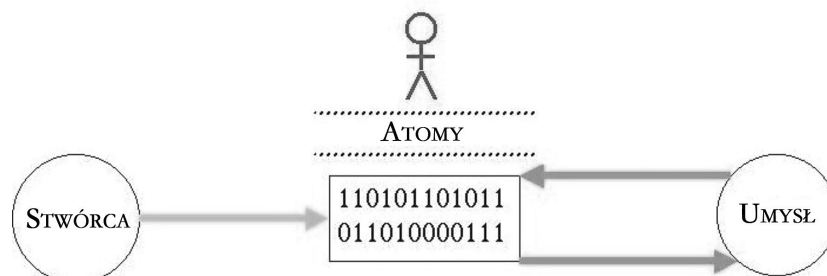
(4) HIPOTEZA METAFIZYCZNA

Możemy teraz połączyć wszystkie powyższe hipotezy. Zajmijmy się najpierw Hipotezą Łączącą, która scali je ze sobą. Zgodnie z nią fizyczna czasoprzestrzeń, wraz z tym, co znajduje się w jej obrębie, jest stworzona przez istoty spoza niej, procesy mikrofizyczne są konstytuowane przez procesy obliczeniowe, a nasze umysły, choć znajdują się na zewnątrz fizycznej czasoprzestrzeni, wchodzą z nią w oddziaływania przyczynowe.

Hipoteza Łącząca, podobnie jak twierdzenia, które się na nią składają, jest spójna i nie możemy jej jednoznacznie wykluczyć. Ponadto nie jest ona sceptyczna, gdyż – mimo że może prowadzić do rekapitulacji niektórych naszych przekonań – na większość z nich nie ma żadnego wpływu.

Zbadajmy teraz Hipotezę Metafizyczną (przez duże M). Podobnie jak Hipoteza Łącząca scala ona Hipotezę Stworzenia, Hipotezę Obliczeniową i Hipotezę Umysł-Ciało. Ponadto wprowadza ona dodatkowe twierdzenie, zgodnie z którym procesy obliczeniowe, leżące u podstaw fizycznej czasoprzestrzeni zostały zaprojektowane przez ich autorów jako komputerowa symulacja świata.

(Hipotezę Metafizyczną można również potraktować jako pogląd, zgodnie z którym procesy obliczeniowe, konstytuujące czasoprzestrzeń fizyczną są częścią szerszej domeny, w obrębie której znajdują się zarówno twórcy symulacji, jak i mój system poznawczy. Takie dopowiedzenie nie jest konieczne, ale dobrze oddaje potoczne rozumienie Hipotezy Matriksa).



Ponieważ Hipoteza Metafizyczna dookreśla relacje pomiędzy częściami Hipotezy Łączącej, można powiedzieć, że jest jej bardziej szczegółową wersją. Ponadto Hipoteza Metafizyczna również jest spójna i nie można jej jednoznacznie wykluczyć. Nie jest to hipoteza sceptyczna: nawet jeśli ją przyjmujemy, większość naszych potocznych przekonań o świecie zewnętrznym nie zmienia się.

4. HIPOTEZA MATRIKSA JAKO HIPOTEZA METAFIZYCZNA

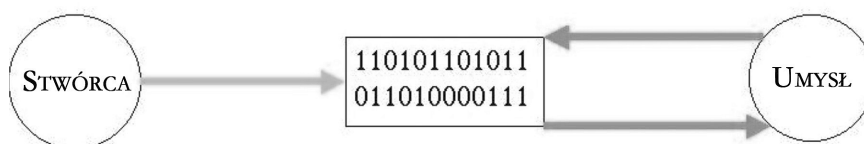
Przypomnijmy, że zgodnie z Hipotezą Matriksa mam (i zawsze miałem) system poznawczy, który odbiera bodźce i wysyła reakcje do sztucznie zaprojektowanej, komputerowej symulacji świata.

Jak będę wskazywał, Hipoteza Matriksa jest równoważna z Hipotezą Metafizyczną, gdyż akceptacja tej pierwszej powinna prowadzić do uznania tej drugiej i *vice versa*. Oznacza to, że obie hipotezy nawzajem się implikują: akceptacja jednej pociąga konieczność uznania drugiej.

Zajmijmy się najpierw przejściem od Hipotezy Metafizycznej do Hipotezy Matriksa. Z Hipotezy Umysł-Ciało wynika, że mam (i zawsze miałem) wyizolowany system poznawczy, który otrzymuje bodźce i wysyła reakcje do fizycznej czasoprzestrzeni. W połączeniu z Hipotezą Obliczeniową oznacza to, że mój system poznawczy wchodzi w interakcje z procesami obliczeniowymi, konstytuującymi fizyczną czasoprzestrzeń. Hipoteza Stworzenia (wraz z resztą Hipotezy Metafizycznej) prowadzi do uznania, że procesy te zostały sztucznie zaprojektowane i to w taki sposób, by symulowały świat. Wynika stąd, że mam (i zawsze miałem) wyizolowany system poznawczy, który otrzymuje bodźce i wysyła reakcje do sztucznie zaprojektowanej komputerowej symulacji świata. I to właśnie głosi Hipoteza Matriksa. A zatem Hipoteza Metafizyczna implikuje Hipotezę Matriksa.

Przejście w drugą stronę jest bardzo podobne. Mówiąc niezbyt ściśle, jeśli przyjmuję Hipotezę Matriksa, akceptuję, że to, co leży u podstaw pozornej rzeczywistości jest określone przez Hipotezę Metafizyczną. Istnieje zatem domena, obejmująca zarówno mój system poznawczy, przyczynowo oddziałujący z komputerową symulacją fizycznej czasoprzestrzeni, jak i inne istoty, autorów tej symulacji. I tego właśnie wymaga Hipoteza Metafizyczna; jeśli zatem ktoś akceptuje taki pogląd, powinien on również przyjąć Hipotezę Stworzenia, Hipotezę Obliczeniową, Hipotezę Umysł-Ciało oraz zachodzące między nimi relacje.

Poniższy rysunek powinien pomóc w zrozumieniu tego schematu. Tak oto, zgodnie z Hipotezą Matriksa, wygląda świat.



Przedstawiony tu poziom fundamentalny jest identyczny z tym, jaki pojawił się przy omawianiu Hipotezy Metafizycznej. Jeżeli zatem przyjmujemy obraz świata zgodny z Hipotezą Matriksa, za poprawny powinniśmy również uznać obraz świata postulowany przez Hipotezą Metafizyczną.

Powyższe twierdzenie nie jest oczywiste. Można na przykład utrzymywać, że Hipoteza Matriksa implikuje istnienie komputerowej symulacji procesów fizycznych, ale (w przeciwieństwie do Hipotezy Metafizycznej) nie implikuje, że jakiegokolwiek procesy fizyczne w ogóle zachodzą. Zarzut ten omówię w punkcie 6, inne zaś obiekcje w punkcie 7. Na razie przyjmuję, że mamy mocne podstawy, by twierdzić, iż Hipoteza Matriksa implikuje Hipotezę Metafizyczną i na odwrót.

5. ŻYCIE W MATRIKSIE

Jeżeli mam rację, oznacza to, że Hipoteza Matriksa nie jest sceptyczna. Nie mogę więc wywodzić z niej, że świat zewnętrzny nie istnieje, że nie mam ciała, że nie istnieją stoły i krzesła albo że nie przebywam w Tucson. Powinienem raczej sądzić, że świat fizyczny jest konstytuowany przez obliczenia, kryjące się pod poziomem mikrofizycznym. A zatem stoły, krzesła

i ciała istnieją, tyle że są złożone z bitów i tego, co je konstytuuje. Świat ten, choć stworzony przez inne istoty, jest całkowicie realny. Mój umysł jest odrębny od procesów fizycznych, ale oddziałuje z nimi przyczynowo. Być może nie został on stworzony, możliwe również, że nie składa się z bitów, lecz wchodzi z nimi w interakcje.

W rezultacie otrzymujemy złożony obraz fundamentalnej natury rzeczywistości. Wydawać się on może dziwny i zaskakujący, ale to wizja pełnoprawnego świata zewnętrznego. Jeżeli jesteśmy w matriksie, to taki po prostu jest świat.

Hipoteza Matriksa może być rozumiana jako charakterystyczny dla ery informacji mit o stworzeniu. Zgodnie z nim, świat fizyczny został stworzony, choć niekoniecznie przez bóstwo. U jego podstaw leży ogromny proces obliczeniowy, którego implementacja była równoznaczna ze stworzeniem. Nasze umysły znajdują się poza strukturą fizyczną i choć wchodzi z nią w związki przyczynowe, posiadają niezależną naturę.

Hipoteza ta generuje pewne kwestie, które pojawiają się również w przypadku bardziej typowych mitów o stworzeniu. Kiedy powstał świat? Mówiąc ściśle, w ogóle nie powstał on w obrębie naszego czasu. Kiedy rozpoczęła się historia? Twórcy mogli wprawdzie uruchomić symulację od roku 4004 p.n.e. (lub od 1999) i spreparować skamieliny, ale bardziej prawdopodobne jest, że początkiem symulacji był Wielki Wybuch, po którym wszystko zaczęło działać samo.

(W *Matriksie* stwórcami są, rzecz jasna, maszyny. To pozwala na interesującą alternatywę w stosunku do teologicznych interpretacji filmu. Zazwyczaj uważa się, że Neo to odpowiednik Chrystusa, Morfeusz to Jan Chrzciciel, a Cypher to Judasz Iskariota. Ale zgodnie z moją tezą bogami Matriksa są maszyny. Kto zatem jest Chrystusem? Oczywiście agent Smith! W końcu to on jest potomkiem bogów, zesłanym, aby ocalić Matriks przed zakusami tych, którzy chcą go zniszczyć. Co więcej, Smith w *sequelu* zmartwychwstaje).

Ponadto pojawiają się tu również kwestie standardowe dla Hipotezy Umysł-Ciało. Kiedy powstają nasze нефизyczne umysły? To zależy, kiedy zamknięte w kadziach systemu poznawcze zostają dołączone do symulacji (możliwe, że albo w chwili poczęcia w obrębie matriksa, albo w czasie narodzin). Czy istnieje życie po śmierci? To zależy, co się dzieje z zamkniętymi w kadziach systemami, kiedy symulacje ich ciał umierają. Jak oddziałują na siebie umysł i ciało? Poprzez związki przyczynowe, zewnętrzne względem fizycznej czasoprzestrzeni.

Nawet jeżeli nie żyjemy w matriksie, moje rozumowanie wciąż dotyczy istot, które w nim są. Odkrywszy swoje położenie, nie powinny one od-

rzucąc zdroworozsądkowych przekonań na temat zewnętrznego świata. Co najwyżej mogą być zmuszone do zrewidowania przekonań odnośnie do fundamentalnej natury ich świata i uznania, że jest on zbudowany z bitów. Istoty te nie żyją jednak w ogromnym złudzeniu, a większość ich potocznych przekonań o rzeczywistości jest poprawna.

Należy jednak poczynić kilka zastrzeżeń. A co z przekonaniem na temat umysłów innych ludzi? Wierzę, że moi przyjaciele są świadomi. Ale czy to prawda, skoro żyję w matriksie? Filmowy Matriks sugerował, że takie przekonanie można utrzymać. Tam jednak wykorzystano wielokadziową wersję matriksa: każdy z moich przyjaciół to zamknięty w kadzi, istniejący w świecie zewnętrznym człowiek, który jest prawdopodobnie tak samo świadomy jak ja. Wyjątek stanowią istoty czysto obliczeniowe, takie jak agent Smith, który nie jest zamknięty w kadzi. To, czy i on jest świadomy, zależy od tego, czy procesy obliczeniowe wystarczają do wygenerowania świadomości. Nie będę teraz rozstrzygał tej kwestii. Możemy ją pominąć poprzez nałożenie na Hipotezę Matriksa warunku, że wszystkie istoty, które postrzegamy, muszą być zamknięte w kadzi. Warunek ten nie jest zresztą konieczny, pytanie bowiem, czy inni ludzie są świadomi, odnosi się również do naszego świata i jest tyleż zasadne, co niezależne od tego, czy znajdujemy się w matriksie, czy nie.

Równie kłopotliwe mogą wydawać się przekonania na temat odległej przeszłości i przyszłości. Pozostaną one jednak niezagrożone, o ile tylko komputerowa symulacja będzie odpowiadać za całą czasoprzestrzeń od Wielkiego Wybuchu aż po kres Wszechświata. Założenie to jest częścią składową Hipotezy Metafizycznej, a tym samym Hipotezy Matriksa, obie bowiem nakładają warunek, by komputerowa symulacja była symulacją całego świata. Możliwe są jednak inne symulacje, które zaczęły się stosunkowo niedawno (tak właśnie jest w *Matriksie*) i takie, które wkrótce się zakończą. W tych wypadkach przekonania zamkniętych w kadzi istot na temat przeszłości lub przyszłości ich świata byłyby fałszywe. Jeżeli jednak symulacja trwa przez całe ich życie, można uznać, że mają one poprawne przekonania na temat obecnego stanu swojego otoczenia.

Być może pod pewnymi względami istoty w matriksie rzeczywiście są zwodzone. Stwórcy mogą kontrolować zdarzenia rozgrywające się w symulacji świata i dowolnie w nie ingerować. (Prawdopodobnie tak właśnie jest w przypadku Matriksa, choć na podstawie filmu trudno ocenić zakres kontroli twórców). Zamknięte w matriksie istoty mogą mieć zatem znacznie mniejszą władzę nad tym, co się dzieje niż im się wydaje. To samo jednak dotyczy sytuacji, w której w niebędący matriksem świat ingeruje Bóg. Hipoteza Matriksa

nie implikuje, że twórcy wpływają na świat, choć jednocześnie nie wyklucza takiej możliwości. W najgorszym razie jest ona pod tym względem nie bardziej sceptyczna niż Hipoteza Stworzenia dla świata niebędącego matriksem.

Mieszkańcy matriksa nie wiedzą również, że rzeczywistość jest dużo większa, niż im się zdaje. Mogą oni mniemać, że istnieje wyłącznie ich fizyczny wszechświat, podczas gdy tak naprawdę wszechświat jest o wiele większy i obejmuje byty, których uwięzione w matriksie istoty nigdy nie będą mogły poznać. Ale problem ten dotyczy również świata, który nie jest matriksem. Kosmologowie poważnie rozważają hipotezę, że za początkiem naszego wszechświata stoi czarna dziura, znajdująca się „o jeden wszechświat wyżej” i że istnieje może rozgałęziona struktura wielu wszechświatów. Nasz świat byłby zatem o wiele większy, niż nam się wydaje, a to, co się w nim znajduje mogłyby pozostać dla nas nieznanne. Nie zmienia to faktu, że świat ten jest całkowicie rzeczywisty.

Co ważne, żadne z tych źródeł sceptycyzmu (odnośnie do innych umysłów, przeszłości i przyszłości, kontroli nad naszym światem czy jego rozmiarów) nie może być źródłem wątpliwości co do prawdziwości naszych przekonań na temat tego, co dane nam zmysłowo. Żadne z nich nie podważa istnienia przedmiotów zewnętrznych, stołów czy krzeseł w taki sposób, w jaki rzekomo ma to robić hipoteza z kadziami. Żadne nie jest też związane wyłącznie ze scenariuszem matriksa. Wątpliwości, czy istnieją inne umysły, przeszłość i przyszłość albo czy mamy kontrolę nad naszym światem są całkowicie niezależne od pytania, czy żyjemy w matriksie. A zatem, w przeciwieństwie do tego, co potocznie się sądzi, Hipoteza Matriksa nie stawia nas w obliczu problemu sceptycyzmu.

Jak zasugerowałem wcześniej, niewykluczone, że naprawdę żyjemy w matriksie. Ta konkluzja może co niektórych zmartwić. Jeżeli jednak mam rację, zmartwienie nie jest tak duże, jak można by mniemać. Nawet jeżeli jesteśmy w matriksie, nasz świat jest nie mniej rzeczywisty, niż wcześniej sądziliśmy. Zaskakująca okazuje się po prostu jego fundamentalna natura.

6. ZARZUT: SYMULACJA TO NIE RZECZYWISTOŚĆ⁷

Zgodnie z najczęściej wysuwany zarzutem symulacja nie jest tym samym, co rzeczywistość. Hipoteza Matriksa mówi jedynie o zachodzeniu

⁷ Rozdział ten ma charakter techniczny i można go bez dużej szkody pominąć.

symulacji procesów fizycznych. Natomiast Hipoteza Metafizyczna implikuje twierdzenie, że procesy te (o których wprost mówi m.in. Hipoteza Obliczeniowa) są rzeczywiste. Hipoteza Matriksa nie może zatem pociągać Hipotezy Metafizycznej – jeżeli jestem w matriksie, żadnych realnych procesów fizycznych nie ma.

Odpowiedź jest następująca: mój argument nie wymaga ogólnego założenia, że symulacja jest tym samym, co rzeczywistość. Całe rozumowanie przebiega nieco inaczej. Zarzut ten pomaga jednak wzmocnić mniej ścisły argument na rzecz twierdzenia, że Hipoteza Matriksa implikuje Hipotezę Metafizyczną.

Ponieważ Hipoteza Obliczeniowa jest spójna, z pewnością możliwe jest, że u podstaw procesów fizycznych leżą procesy obliczeniowe, zaimplementowane w dalszych procesach. Jakiś rodzaj systemu obliczeniowego mógłby zatem stanowić podstawę rzeczywistości. Oponent będzie jednak twierdził, że nie wszystkie systemy obliczeniowe są sobie równe. Nawet jeżeli niektóre z nich mogłyby być podstawą rzeczywistych procesów fizycznych, nie oznacza to, że wszystkie takie są, część z nich bowiem to wyłącznie symulacje. W takim wypadku Hipoteza Matriksa nie odnosi się do podstaw rzeczywistości.

Aby oddalić ten zarzut, odwołajmy się do dwóch zasad. Zgodnie z pierwszą każde abstrakcyjne obliczenie, które może być wykorzystane do symulowania fizycznej czasoprzestrzeni, mogłoby stanowić podstawę prawdziwych procesów fizycznych. Zgodnie z drugą, jeżeli abstrakcyjne obliczenie mogłoby stanowić podstawę procesów fizycznych, dokładny sposób jego implementacji jest nieważny dla faktu, czy istotnie odgrywa ono taką rolę. W szczególności nie ma znaczenia to, czy implementacja została zaprojektowana jako symulacja. Z tego bezpośrednio wynika nasz wniosek.

Co do pierwszej zasady: potraktujmy abstrakcyjne obliczenia czysto formalnie, pomijając sposób ich implementacji. Aby abstrakcyjne obliczenie mogło być uznane za symulację fizycznej rzeczywistości, musi ono posiadać elementy obliczeniowe odpowiadające każdej części występującej w rzeczywistości (jak również polu, fali i innym własnościom fundamentalnym), dynamicznie rozwijające się w sposób, który odpowiada zachowaniu cząstki. Mamy wtedy pewność, że struktura obliczenia będzie wystarczająco bogata przyczynowo, by mogła ona stanowić podstawę fizyki naszego świata. Tak działa każde obliczenie, o ile jest wystarczająco dokładne, by odpowiadać szczegółowością procesom fizycznym.

Co do drugiej zasady: jeżeli dysponujemy abstrakcyjnym obliczeniem, które mogłoby stanowić podstawę fizycznej rzeczywistości, sposób jego im-

plementacji nie ma żadnego znaczenia. Wyobraźmy sobie, że odkrywamy poziom obliczeniowy, który stanowi podstawę poziomemu atomowego czy elektronowego. Możliwe, że poziom obliczeniowy sam zaimplementowany jest w procesach bardziej fundamentalnych. Jest wiele teorii wyjaśniających naturę podstawowych procesów, ale żadna z nich nie jest uprzywilejowana ani nie zmusza nas do odrzucenia hipotezy, że poziom obliczeniowy konstytuuje procesy fizyczne. Oznacza to, że Hipoteza Obliczeniowa jest implementacyjnie niezależna: jeżeli dysponujemy właściwym rodzajem abstrakcyjnego obliczenia, sposób jego implementacji nie odgrywa żadnej roli.

W szczególności znaczenia nie ma, czy zaimplementowane procesy stworzono sztucznie i czy zamierzono je jako symulację. Uwzględniamy nie ich pochodzenie, lecz wyłącznie wewnętrzną naturę, w której istotne jest to, by procesy były zorganizowane tak, aby implementować właściwy rodzaj obliczenia. Fakt, że implementację zaplanowano jako symulację, nie ma zatem żadnego wpływu na jej zdolności do konstytuowania rzeczywistości fizycznej.

Istnieje jeszcze jedno ograniczenie implementacji: procesy muszą się odpowiednio łączyć z naszymi przeżyciami. Kiedy doświadczamy przedmiotu, procesy leżące u podstaw jego symulacji muszą we właściwy sposób łączyć się przyczynowo z naszymi przeżyciami. W przeciwnym razie nie mielibyśmy podstaw, by uważać procesy obliczeniowe za rzeczywiste źródło postrzeganych przez nas procesów fizycznych. W przypadku wyizolowanej symulacji komputerowej, do której nikt nie jest podłączony, powiedzieliśmy, że to po prostu symulacja. Ale odpowiednie jej podłączenie do naszych przeżyć zmysłowych to istotny element najbardziej naturalnego rozumienia Hipotezy Matriksa. Nie generuje ona zatem w tym miejscu żadnych trudności.

Podsumowując: procesy obliczeniowe mogą być podstawą fizycznej rzeczywistości, każde abstrakcyjne obliczenie, które kwalifikuje się jako symulacja fizycznej rzeczywistości, może odgrywać tę rolę, a każda implementacja tego obliczenia może konstytuować fizyczną rzeczywistość, o ile symulacja ta łączy się odpowiednio z naszym doświadczeniem. Hipoteza Matriksa gwarantuje, że dysponujemy odpowiednim rodzajem abstrakcyjnego obliczenia i że połączenie jest właściwe. Wynika z niej zatem, że Hipoteza Obliczeniowa jest poprawna, a komputerowa symulacja konstytuuje prawdziwe procesy fizyczne.

7. INNE ZARZUTY

Kiedy patrzymy z zewnątrz na mózg w kadzi, trudno pozbyć się wrażenia, że jest on zwodzony. To właśnie przeświadczenie generuje pewne powiązane ze sobą zarzuty. Nie są one skierowane bezpośrednio przeciwko mojej argumentacji, ale kwestionują jej wniosek.



ZARZUT 1. Mózg w kadzi myśli, że spaceruje w promieniach słońca, podczas gdy tak naprawdę przebywa samotnie w ciemnym pokoju. To oczywiste, że jest zwodzony!

ODPOWIEDŹ. Mózg istotnie znajduje się sam w ciemnym pokoju, ale to nie znaczy, że podobnie jest z osobą. Posłużmy się pewną analogią. Załóżmy, że Kartezjusz ma rację: nasze pozbawione ciał, zbudowane z ekto-plazmy umysły znajdują się poza czasoprzestrzenią. Gdyby jakiś anioł spojrział na mój ekto-plazmatyczny umysł w chwili, gdy myślę „jestem na zewnątrz w promieniach słońca”, stwierdziłby, że w rzeczywistości nie jestem wystawiony na działanie słońca. Czy to znaczy, że się mylę? Raczej nie: mogę przebywać na zewnątrz w promieniach słońca, nawet jeżeli nie może się tam znajdować mój ekto-plazmatyczny umysł. Ów anioł nie miałby zatem racji, sądząc, że moje przekonanie jest błędne. Analogicznie, nie powinniśmy wnioskować, że przekonanie istoty zamkniętej w kadzi byłoby nieprawdziwe. Jeżeli jest ona zwodzona, to nie bardziej niż kartezjański umysł.

Wynika stąd, że większość naszych przekonań może być prawdziwa bez względu na bezpośrednie otoczenie naszych umysłów. Ważne są jedynie procesy, do których za pomocą zmysłowych bodźców i ruchowych reakcji podłączony jest nasz umysł. Wystarczy to zauważyć, a zarzut upada.

ZARZUT 2. Zamknięta w kadzi istota może uważać, że jest w Tucson, a tak naprawdę znajduje w Nowym Jorku i nigdy nie przebywała nawet w pobliżu Tucson. To oczywiste, że jej przekonanie jest efektem zwodzenia.

ODPOWIEDŹ. Pojęcie „Tucson”, którym posługuje się zamknięty w kadzi mózg, nie odnosi się do tego, co my określamy tą nazwą. Oznacza ono coś całkowicie innego, co nazwać możemy Tucson* lub „wirtualnym Tucson”, będącym pewnego rodzaju „wirtualnym miejscem” (więcej o tym za chwilę). Kiedy zamknięta w kadzi istota mówi: „jestem w Tucson”, ma ona na myśli, że przebywa w Tucson*, co jest zresztą zgodne z prawdą. Ponieważ Tucson to nie Tucson*, fakt, że nigdy nie była ona w Tucson pozostaje bez wpływu na prawdziwość jej przekonania.

Weźmy taką oto luźną analogię: patrzę na mojego kolegę Tadeusza i myślę: „oto Tadeusz”. W innym miejscu na świecie duplikat mnie patrzy na duplikat Tadeusza. I choć myśli: „oto Tadeusz”, nie spogląda wcale na prawdziwego Tadeusza. Czy to przekonanie jest fałszywe? Wydaje się, że nie: pojęcie „Tadeusza”, którym posługuje się mój duplikat, odnosi się nie do Tadeusza, lecz do jego klonu, czyli Tadeusza*. Ponieważ mój duplikat naprawdę patrzy na Tadeusza*, jego przekonanie jest prawdziwe. To samo dzieje się w przedstawionym wyżej przypadku.

ZARZUT 3. Zanim Neo opuścił Matriks, był przekonany, że ma włosy. W rzeczywistości jednak ich nie posiada (ciało w kadzi jest ich pozbawione). To przekonanie jest zatem ewidentnie efektem iluzji.

ODPOWIEDŹ. Ten przypadek przypomina poprzedni zarzut. Pojęcie „włosy”, którym posługuje się Neo, nie odnosi się do prawdziwych włosów, ale do czegoś, co moglibyśmy nazwać włosami* („wirtualnymi włosami”). Fakt, że Neo nie ma rzeczywistych włosów, jest zatem bez znaczenia dla prawdziwości jego przekonania. Tak naprawdę Neo posiada wirtualne włosy, a więc ma rację. Podobnie jest w przypadku, gdy w filmie dziecko mówi do Neo: „tu nie ma łyżki”; pojęcie „łyżka” odnosi się do wirtualnej łyżki, która tak naprawdę istnieje. A zatem dziecko się myli.

ZARZUT 4. Do jakiego rodzaju przedmiotów odnoszą się zamknięte w kadzi mózgi? Czym są wirtualne włosy, wirtualne Tucson i tak dalej?

ODPOWIEDŹ. Każdy z tych bytów konstituowany jest przez procesy obliczeniowe. Jeśli byłbym zamknięty w kadzi, wszystkie przedmioty, do których się odnoszę (włosy, Tucson itd.), zbudowane byłyby z bitów. Jeśli zaś w kadzi zamknięta byłaby inna istota podłączona do symulacji w moim komputerze, obiekty, do których się ona odnosi, byłyby konstituowane przez zestawy bitów wewnątrz mojego komputera. Tego rodzaju przedmioty

(włosy*, Tucson* itd.) możemy nazwać wirtualnymi. Choć wirtualne dłonie nie są dłońmi (założywszy, że nie jestem zamknięty w kadzi), wciąż istnieją one w komputerze. Wirtualne Tucson to nie Tucson, ale może ono istnieć we wnętrzu komputera.

ZARZUT 5. Przed chwilą stwierdziłeś, że wirtualne dłonie to nie prawdziwe dłonie. Czy nie znaczy to, że jeżeli jesteśmy w matriksie, to nie mamy prawdziwych dłoni?

ODPOWIEDŹ. Nie. Jeśli nie jesteśmy w matriksie, ale przebywa w nim ktoś inny, powinniśmy powiedzieć, że jego termin „dłonie”, w przeciwieństwie do naszego, odnosi się do wirtualnych dłoni. W takiej sytuacji nasze dłonie nie są wirtualnymi dłońmi. Ale jeśli to my jesteśmy w matriksie, wtedy nasz termin „dłonie” odnosi się do czegoś, co zrobione jest z bitów – wirtualnych dłoni, a przynajmniej tego, co byłoby za nie uważane przez ludzi ze świata „o piętro wyższego”. Jeżeli zatem jesteśmy w matriksie, prawdziwe dłonie zrobione są z bitów. W zależności od tego, czy przyjmujemy perspektywę z zewnątrz czy z wewnątrz matriksa, rzeczy wyglądają inaczej, a nasze słowa odnoszą się do innych przedmiotów,

Taka zmiana perspektywy zdarza się często, gdy myślimy o matriksie. Jeżeli z perspektywy pierwszoosobowej założymy, że to my jesteśmy w matriksie, wtedy realne rzeczy w naszym świecie składają się z bitów, choć inaczej może być ze światem „o piętro wyżej”. Jeżeli z perspektywy trzecioosobowej przyjmiemy, że w matriksie jesteśmy nie my, ale ktoś inny, to realne rzeczy w naszym świecie nie są zrobione z bitów, w przeciwieństwie do świata „o piętro niżej”. W pierwszym wypadku do bitów obliczeniowych odnoszą się nasze terminy, w drugim – terminy zamkniętych w kadziach istot.

ZARZUT 6. Ale który zestaw bitów odpowiada za dany przedmiot wirtualny? To przecież jasne, że nie da się tego określić.

ODPOWIEDŹ. Takie pytanie przypomina próbę docieczenia, która część kwantowej funkcji falowej jest tym krzesłem albo Uniwersytetem w Arizonie. Choć przedmioty te są ostatecznie konstytuowane przez leżące u ich podstaw kwantowe funkcje falowe, być może nie ma żadnej konkretnej części takiej funkcji, którą moglibyśmy wskazać i stwierdzić, że „jest” krzesłem albo uniwersytetem. Krzesło i uniwersytet istnieją na wyższym poziomie. Analogicznie, jeżeli bylibyśmy zamknięci w kadzi, wskazanie konkretnego zestawu bitów, będącego na mikropoziomie obliczeniowym krzesłem czy uniwersytetem, mogłoby się okazać niemożliwe. Przedmioty te po prostu znajdują się na wyższym poziomie. Jeżeli natomiast zamknięty

w kadzi jest ktoś inny, nie istnieją być może określone zestawy bitów, będące w komputerowej symulacji przedmiotami, do których istota ta się odnosi. Tak jak krzesło istnieje bez określania konkretnej części funkcji falowej, tak wirtualne krzesło może istnieć bez określania konkretnego zestawu bitów.

ZARZUT 7. Zamknięta w kadzi istota myśli, że wykonuje działania albo że ma przyjaciół. Czy te przekonania są poprawne?

ODPOWIEDŹ. Jedną z możliwości jest stwierdzenie, że istota ta wykonuje działania* i że ma przyjaciół*. Wydaje mi się jednak, że z różnych względów błędne jest założenie, że słowa takie jak „działanie” czy „przyjaciel” mogą zmieniać znaczenie z taką łatwością jak „Tucson” czy „włosy”. Uważam raczej, że można zgodnie z prawdą stwierdzić (w naszym języku), że zamknięty w kadzi byt wykonuje działania i że ma przyjaciół. Mówiąc ściślej: wykonuje on działania w swoim środowisku, przy czym jego wirtualne środowisko nie jest naszym środowiskiem. Tę wirtualną przestrzeń zamieszkują również jego przyjaciele (przy założeniu, że matriks obejmuje wiele kadzi albo że do zaistnienia świadomości wystarczają procesy obliczeniowe). W tych jednak przypadkach zamknięta w kadzi istota nie myli się.

ZARZUT 8. Zostawmy na boku te kwestie techniczne. To oczywiste, że jeżeli żyjemy w matriksie, świat w niczym nie przypomina tego, co o nim sądzą!

ODPOWIEDŹ. Nie zgadzam się. Nawet jeśli jesteśmy w matriksie, wciąż istnieją ludzie, mecze piłki nożnej i cząsteczki rozmieszczone w czasoprzestrzeni w taki sposób, w jaki nam się wydaje. Świat ma po prostu jakąś głębszą naturę, która wykracza poza naszą początkową koncepcję. Rzeczy w świecie są realizowane obliczeniowo w sposób, którego początkowo nie moglibyśmy sobie nawet wyobrazić. Ale nie popadamy w żadną sprzeczność z naszymi potocznymi przekonaniem. Musimy co najwyżej odrzucić kilka bardziej abstrakcyjnych twierdzeń metafizycznych. Dokładnie to samo dzieje się jednak, gdy uznamy poprawność mechaniki kwantowej czy teorii względności.

Jeżeli żyjemy w matriksie, większość naszych przekonań jest prawdziwa, zarazem jednak posiadamy istotne luki w wiedzy; nie wiemy na przykład, że wszechświat jest realizowany obliczeniowo. Ale tego właśnie powinniśmy oczekiwać. Nawet jeśli nie jesteśmy w matriksie, możemy mieć poważne braki w rozumieniu fundamentalnej natury rzeczywistości. Nie jesteśmy wszechwiedzący, a nasza wiedza o świecie jest w najlepszym razie częściowa. To całkowicie zwyczajny stan istot zamieszkujących jakiegokolwiek świat.

8. ALTERNATYWNE HIPOTEZY SCEPTYCKIE

Hipoteza Matriksa to tylko jeden z przykładów tradycyjnej hipotezy „sceptycznej”. Inne jej wersje nie są tak ewidentnie sceptyczne, ale stosuje się do nich podobna kontrargumentacja. Prawdziwość większości z nich nie prowadzi do zakwestionowania naszych empirycznych przekonań na temat fizycznego świata, stąd nie mogą one służyć jako argumenty na rzecz sceptycyzmu globalnego. Mają one co najwyżej charakter częściowo sceptyki: podważają niektóre z naszych przekonań, lecz pozostawiając bez szwanku inne.

HIPOTEZA NOWEGO MATRIKSA. Zostałem niedawno stworzony, obdarzony fałszywymi wspomnieniami, a następnie umieszczony w nowo utworzonym matriksie.

A co jeśli zarówno matriks, jak i ja istniejemy od niedawna? Tak brzmi obliczeniowa wersja zaproponowanej przez Bertranda Russella Hipotezy Niedawnego Stworzenia, zgodnie z którą ja oraz świat fizyczny zostaliśmy stworzeni niedawno (wraz z – odpowiednio – wspomnieniami i skamielinami). Choć świat zewnętrzny, który postrzegam, naprawdę istnieje, a większość moich przekonań na temat jego obecnego stanu jest prawdziwa, fałszywa okazuje się znaczna część przekonań o przeszłości. Podobnie jest z połączeniem Hipotezy Niedawnego Stworzenia oraz Hipotezy Metafizycznej, czyli Hipotezą Nowego Matriksa; nie ma ona charakteru globalnie sceptycznego (sceptycyzm ten jest częściowy i dotyczy wyłącznie przekonań o przeszłości).

HIPOTEZA NIEDAWNEGO MATRIKSA. Choć przez większość życia nie byłem zamknięty w kadzi, niedawno podłączono mnie do matriksa.

Wydaje się, że jeżeli niedawno podłączono mnie do matriksa, a ja nawet tego nie zauważyłem, większość przekonań na temat mojego obecnego środowiska jest fałszywa. Załóżmy, że ktoś umieścił mnie wczoraj w symulacji, w której lecę do Las Vegas, a następnie idę do kasyna. Mimo że jestem przekonany, że przebywam teraz w kasynie w Las Vegas, przekonanie to jest fałszywe: tak naprawdę znajduję się w laboratorium w Tucson.

Rezultat jest zatem inny niż w przypadku długoterminowego matriksa. Różnica bierze się stąd, że moje pojęcie zewnętrznej rzeczywistości odnosi się do tej rzeczywistości, w której spędziłem większość życia. Jeśli byłbym zamknięty w kadzi od urodzenia, moje pojęcie odnosiłoby się do rzeczywistości konstytuowanej obliczeniowo. Jeżeli jednak zostałbym zamknięty w kadzi wczoraj, odnosiłoby się ono do rzeczywistości zewnętrznej. Stąd,

jeśli sądziłbym, że przebywam w Las Vegas, miałbym na myśli zewnętrzne Las Vegas, co powodowałoby, że byłbym w błędzie.

Ale i ta sytuacja nie podważa wszystkich moich przekonań o zewnętrznym świecie. Uważam, że urodziłem się w Sydney albo że w oceanach jest woda, a te i podobne przekonania pozostają poprawne. Fałszywe są wyłącznie przekonania pozyskane ostatnio, pochodzące z percepcji symulowanego środowiska. Hipoteza Niedawnego Matriksa jest zatem częściowo sceptyczna: jej dopuszczalność może spowodować, że zwątpimy w część naszych przekonań empirycznych, ale nie we wszystkie.

Mimo podobieństw, Hipoteza Niedawnego Matriksa i Hipoteza Nowego Matriksa dają przeciwstawny rezultat: pierwsza uzasadnia prawdziwość przekonań o przeszłości, ale wskazuje fałszywość przekonań o teraźniejszości, podczas gdy druga uzasadnia fałszywość przekonań o przeszłości, lecz prowadzi do prawdziwości przekonań o teraźniejszości. Różnica bierze się z faktu, że zgodnie z Hipotezą Niedawnego Matriksa, a w przeciwieństwie do Hipotezy Nowego Matriksa, naprawdę istniałem w przeszłości; i to właśnie w przeszłych doświadczeniach zakorzenione są moje myśli.

HIPOTEZA LOKALNEGO MATRIKSA. Zostałem podłączony do komputerowej symulacji lokalnego fragmentu świata.

Przyjmijmy, że komputer symuluje mały odcinek świata, którego granic z różnych względów nie można opuścić. Dla przykładu: w filmie *Trzynaste piętro* symulowana jest wyłącznie Kalifornia, ten zaś, kto próbuje dostać się do Newady, napotyka blokady drogowe z napisem „Zamknięte na czas remontu” (a w oddali wznoszą się bładozielone, elektroniczne góry!). Nie jest to najlepszy sposób tworzenia matriksa, gdyż jego mieszkańcy prędzej czy później odkryją ograniczenia świata, w którym żyją.

Przypadek ten odpowiada Hipotezie Lokalnego Stworzenia, zgodnie z którą stwórcy skonstruowali tylko małą część fizycznego świata. W takim wypadku prawdziwe są nasze przekonania o pobliskich przedmiotach, fałszywe zaś te, które dotyczą czegoś położonego dalej. To samo dotyczy Hipotezy Lokalnego Matriksa, która może być traktowana jako połączenie Hipotezy Metafizycznej i Hipotezy Lokalnego Stworzenia.

HIPOTEZA ROZSZERZALNEGO LOKALNEGO MATRIKSA. Jestem podłączony do komputerowej symulacji lokalnego fragmentu świata, który, w zależności od działań podmiotu, może się rozszerzać.

Przy takim rozwiązaniu unikamy oczywistych trudności, z jakimi boryka się Hipoteza Lokalnego Matriksa. Symulacja dotyczy tylko środowiska lokalnego, lecz jest rozszerzana, jeżeli zachodzi taka potrzeba. Dzięki temu

stwórcy mogą na przykład koncentrować się teraz na symulacji jednego pokoju mojego domu w Tucson; jeżeli przeniosę się do innego pokoju albo polecę do innego miasta, będą oni symulować każde z tych miejsc. Oczywiście muszą oni dopilnować, by przy ich tworzeniu uwzględnić moje wspomnienia i przekonania oraz zmiany, jakim owo miejsce podlega w międzyczasie. To samo dotyczy napotkanych znajomych lub ludzi, których znam ze słyszenia. Najlepiej założyć, że stwórcy dysponują bazą danych, zawierającą informacje o tym, co w świecie zostało już skonstruowane, i aktualizują ją lub wprowadzają do niej nowe dane, kiedy zachodzi taka potrzeba.

Tego rodzaju symulacja nie przypomina zwyczajnego matriksa, w którym stworzono od razu cały świat. W przypadku tego drugiego koszty początkowe są duże, ale kiedy symulacja zacznie działać, nie potrzebuje dalszego nadzoru. Natomiast rozszerzalny lokalny matriks wymaga symulacji w czasie rzeczywistym. Choć koszty początkowe są dużo mniejsze, wraz z rozwojem symulacja wymaga coraz większych nakładów pracy i kreatywności.

Jeżeli chodzi o koncepcję świata rzeczywistego, hipoteza ta jest podobna do Hipotezy Rozszerzalnego Lokalnego Stworzenia, gdzie stwórcy symulują tylko lokalne środowisko fizyczne i rozszerzają je w miarę potrzeby. A zatem zewnętrzna rzeczywistość istnieje, wiele z naszych przekonań na temat bliskich rzeczy jest prawdziwych, ale przekonania o tym, co dalej położone wciąż są fałszywe. Jeśli połączymy tę hipotezę z Hipotezą Metafizyczną, otrzymamy Hipotezę Rozszerzalnego Lokalnego Matriksa. Jeżeli żyjemy w takim matriksie, zewnętrzna rzeczywistość istnieje, ale znaczna jej część jest inna, niż nam się wydaje. Oczywiście jeżeli będę poruszał się we właściwym kierunku, jakiś jej fragment może rzeczywiście zaistnieć!

Sytuacja ta przypomina *Truman Show*. Tytułowy bohater żyje w sztucznym środowisku, wypełnionym rekwizytami i aktorami, którzy zachowują się normalnie w obecności Trumana, ale kiedy go nie ma – wszystko się zmienia. Wiele przekonań Trumana o obecnym środowisku, na przykład co do stołów i krzeseł, jest prawdziwych. Myli się on jednak co do rzeczy spoza tego środowiska, tzn. położonych dalej od jego domu.

Wydawać się może, że choć *Truman Show* przedstawia nieprzyjemny scenariusz sceptyczny, *Matrix* jest o wiele gorszy. Jeżeli mam rację, jest dokładnie na odwrót. Jeśli jestem w matriksie, większość moich przekonań na temat zewnętrznego świata jest prawdziwa. Jeżeli znajduję się w czymś na podobieństwo *Truman Show*, znaczna część moich przekonań jest fałszywa. Po namyśle uważam, że to prawidłowa konkluzja. Jeśli odkrylibyśmy, że żyjemy w matriksie, byłibyśmy zaskoczeni, ale moglibyśmy szybko

do tego przywyknąć. Jeżeli okazałoby się, że żyjemy w programie typu Truman Show, najpewniej byśmy zwariowali.

HIPOTEZA MAKROSKOPOWEGO MATRIKSA. Jestem podłączony do komputerowej symulacji makroskopowych procesów fizycznych, w której nie generuje się szczegółowego poziomu mikroskopowego.

Wyobraźmy sobie matriks, który dla uproszczenia pozbawiono fizyki niskiego poziomu. Jego stwórcy symulują wyłącznie przedmioty makroskopowe oraz ich własności: np. stół o takim a takim kształcie, położeniu czy kolorze oraz leżącą na blacie książkę, obdarzoną określonymi własnościami. Choć autorzy symulacji musieliby dopilnować, by przedmioty zachowywały się zgodnie z prawami fizyki oraz uwzględnić możliwość dokonywania pomiarów mikrofizycznych, możemy sobie bez trudu wyobrazić taki scenariusz.

Hipoteza ta jest analogiczna do Hipotezy Makroskopowego Świata, zgodnie z którą procesy mikrofizyczne nie istnieją, a fundamentalny charakter mają zastępujące je makroskopowe przedmioty fizyczne, posiadające własności takie jak kształt, kolor czy położenie. Taka wizja świata jest spójna i nie ma charakteru globalnie sceptycznego, choć może prowadzić do falsyfikacji naukowych twierdzeń o niższych poziomach rzeczywistości. Koniunkcja tej hipotezy i Hipotezy Metafizycznej daje Hipotezę Makroskopowego Matriksa; jako taka nie jest ona sceptyczna.

Powyższe hipotezy można ze sobą dowolnie łączyć, uzyskując na przykład Hipotezę Nowego Lokalnego Makroskopowego Matriksa. Wszystkie one są analogiczne do odpowiednich hipotez na temat fizycznego świata. Bez wyjątku da się je zatem pogodzić z istnieniem rzeczywistości fizycznej i żadna nie ma charakteru sceptycznego.

HIPOTEZA BOGA. Rzeczywistość fizyczna jest reprezentowana w umyśle Boga, od którego zależne są nasze myśli i postrzeżenia.

Hipotezę tę, stanowiącą odpowiedź na pytanie, jaki może być świat, zaproponował George Berkeley, traktując ją zarazem jako metafizyczny pogląd na naturę rzeczywistości. Większość filozofów traktowała takie rozwiązanie jako sceptyczne. Jeżeli jednak mam rację, to Berkeley jest bliżej prawdy. Hipotezę Boga można potraktować jako odmianę Hipotezy Matriksa, z tą jednak różnicą, że symulacja świata jest zaimplementowana w umyśle Boga. Powinniśmy zatem stwierdzić, że procesy fizyczne zachodzą naprawdę, na poziomie jednak bardziej fundamentalnym są one konstytuowane przez procesy w Boskim umyśle⁸.

⁸ Chalmers chce najwyraźniej powiedzieć, że zgodnie z omawianą hipotezą postrzegamy idee w umyśle Boga; interpretację taką sugeruje wykorzystanie pojęć implementacji, a zwłaszcza

HIPOTEZA ZŁEGO GENIUSZA. Mój umysł pozbawiony jest ciała, a zły geniusz dostarcza mu bodźców zmysłowych, aby wygenerować pozór istnienia zewnętrznego świata.

Jest to klasyczny sceptyczny argument Kartezjusza. Cóż możemy o nim powiedzieć? To zależy, jak działa zły geniusz. Jeśli symuluje on cały świat w swojej głowie, określając jakie bodźce powinienem otrzymać, mamy do czynienia z odmianą Hipotezy Boga. Powinniśmy zatem stwierdzić, że rzeczywistość fizyczna istnieje i jest konstytuowana przez procesy zachodzące w głowie geniusza. Jeśli symuluje on jedynie niewielki fragment fizycznego świata, wystarczająco duży jednak, by dostarczyć mi spójnych bodźców, wtedy hipoteza ta jest analogiczna do Hipotezy Lokalnego Matriksa (w wersji stałej lub rozszerzalnej). W takim wypadku stwierdzamy, że istnieje lokalna część zewnętrznego świata. Jeżeli zły geniusz nie zadaje sobie trudu, by symulować poziom mikrofizyczny, mamy do czynienia z poglądem podobnym do Hipotezy Makroskopowego Matriksa. W tej sytuacji, choć makroskopowe, lokalne przedmioty zewnętrzne istnieją, nasze przekonania o ich mikroskopowej naturze są błędne.

Hipoteza Złego Geniusza często traktowana jest jako zarazem sceptyczna i globalna. Ale jeżeli moja argumentacja jest poprawna, takie rozumienie jest błędne. Nawet jeśli hipoteza ta okazałaby się prawdziwa, wciąż realnie istniałaby część zewnętrznej, postrzeganej przez nas rzeczywistości, choć niektóre przekonania na jej temat, w zależności od szczegółowych rozstrzygnięć, mogą okazać się fałszywe. Zewnętrzna rzeczywistość ma po prostu leżącą u jej podstaw naturę, która jest nieco inna niż mogło nam się wydawać.

HIPOTEZA SNU. Śnię – i od zawsze śniłem.

Kartezjusz pytał: skąd możesz wiedzieć, że aktualnie nie śnisz? Morfeusz zadaje podobne pytanie: „Czy kiedykolwiek miałeś sen tak realistyczny, iż byłeś pewny, że to rzeczywistość? A co jeśli nie byłbyś w stanie się z niego wybudzić? Na jakiej podstawie odróżniłbyś świat snu i świat rzeczywisty?”.

konstytucji symulowanych procesów przez umysł Boga. Jeżeli tak, to pogląd taki przyjmował N. Malebranche, ale Berkeley stanowczo go odrzucał. Angielski biskup pisze: „Nie wyobrażam ja sobie, ażebyśmy widzieli Boga (jak chcą mieć niektórzy) prostym i bezpośrednim wejrzaniem albo żebyśmy rzeczy materialne widzieli nie przez nie same, lecz widzieli jedynie to, co jest ich obrazem, co je reprezentuje w istocie Boga – która to teoria, przyznaję się, jest dla mnie niepojęta” (George BERKELEY, *Rzecz o zasadach poznania ludzkiego*, przeł. Feliks Jezierski (Warszawa: Hachette, 2009), 320). Fakt, że dla Berkeleya idee, które postrzegamy, nie są tożsame z ideami, które znajdują się w Bożym umyśle, podkreśla również Frederick COPLESTON, *Historia filozofii*, t. V: *od Hobbesa do Hume’a*, przeł. Jarosław Pasek, Joanna Pasek, Paweł Józefowicz (Warszawa: IW PAX, 1997), 372. (Przyp. tłum.).

Hipoteza, że obecnie śnię podobna jest do Hipotezy Niedawnego Matriksa. Nie mogę jej jednoznacznie wykluczyć, zaś jeśli jest ona prawdziwa, to większość moich przekonań co do obecnego środowiska jest niepoprawna. Wciąż jednak prawdziwe są moje przekonania dotyczące przeszłości.

A co jeśli śnię od zawsze? Co jeśli wszystkie moje bodźce wrazeniowe są generowane przez mój system poznawczy, a ja nawet nie zdaję sobie z tego sprawy? Myślę, że scenariusz taki jest podobny do Hipotezy Złego Geniusza, w której rolę zwodziciela odgrywa mój własny system poznawczy! Jeśli system, który generuje sen, symuluje również całą czasoprzestrzeń, mielibyśmy do czynienia z czymś podobnym do oryginalnej Hipotezy Matriksa. Jeśli zaś modeluje tylko moje lokalne środowisko albo pewne procesy makroskopowe, przypadek ten upodabnia się do bardziej lokalnych wersji Hipotezy Złego Geniusza. We wszystkich tych przypadkach powinniśmy twierdzić, że przedmioty, które teraz postrzegam, istnieją naprawdę (choć inaczej może być z rzeczami położonymi dalej). Niektóre z nich są po prostu konstytuowane przez moje własne procesy poznawcze.

HIPOTEZA CHAOSU. Nie otrzymuję bodźców ze świata. Moje doświadczenie jest losowe i nie ma przyczyny. Moje poznanie jest regularne i ustrukturyzowane w taki sposób, w jaki mi się jawi za sprawą ogromnego zbiegu okoliczności.

Hipoteza Chaosu jest bardzo mało prawdopodobna, o wiele mniej niż jakakolwiek z rozważanych wcześniej opcji. Wciąż jednak nie można jej wykluczyć, nawet jeśli jej możliwość jest bliska zeru. Czy zachodzące w zewnętrznym świecie procesy fizyczne istnieją, jeżeli jestem chaotycznym mózgiem uwięzionym w kadzi? Myślę, że odpowiedź jest przecząca. Moje doświadczenie przedmiotów zewnętrznych nie ma żadnej przyczyny, a jego źródło jest inne niż przyczyna pojęć tych przedmiotów. Tak naprawdę moje doświadczenie w ogóle nie jest powodowane przez zewnętrzną rzeczywistość. Hipoteza ta jest zatem w ścisłym sensie sceptyczna: jej przyjęcie skutkowałoby odrzuceniem większości naszych przekonań na temat zewnętrznego świata.

Hipoteza Chaosu jest jak dotąd jedynym ewidentnym przypadkiem hipotezy globalnie sceptycznej. W przeciwieństwie do innych hipotez jej przyjęcie powodowałoby konieczność odrzucenia wszystkich naszych kluczowych przekonań na temat świata zewnętrznego. Z czego bierze się ta różnica?

Możliwe, że Hipotezę Chaosu wyróżnia całkowity brak przyczynowego wyjaśnienia zarówno naszego doświadczenia, jak i jego regularności. We wszystkich poprzednich przypadkach istniało wyjaśnienie tych faktów, choć nie takie,

jakiego byśmy oczekiwali. Można stąd wnosić, że dana hipoteza nie będzie miała charakteru globalnie sceptycznego, o ile będzie dopuszczała jakiegokolwiek racjonalne wytłumaczenie regularności naszego doświadczenia.

Jeżeli założymy, że takie wyjaśnienie istnieje, możemy bezpiecznie orzec, że niektóre przekonania na temat zewnętrznego świata są poprawne. To nie-dużo, ale zawsze coś!

9. UWAGI FILOZOFICZNE

Ponieważ artykuł ten został napisany tak, by był dostępny szerokiej publiczności, celowo unikałem szczegółów technicznych czy odwołań do literatury przedmiotu. W tym miejscu spróbuję uzupełnić te braki. Czytelnicy bez przygotowania filozoficznego mogą pominąć tę część lub tylko ją przekartkować.

UWAGA 1. Jak argumentował Hilary PUTNAM (1981), możemy *a priori* wykluczyć, iż jestem (i zawsze byłem) mózgiem w kadzi. Słowo „mózg” odnosi się bowiem do przedmiotu w świecie postrzeganym, a nie do czegoś w świecie „zewnętrznym”, w którym musiałaby istnieć kadź. Aby moja hipoteza – „jestem mózgiem w kadzi” – mogła być prawdziwa, muszę być mózgiem takiego rodzaju, jaki istnieje w zmysłowo danym mi świecie, co jest niemożliwe. Hipoteza ta musi być zatem fałszywa.

Weźmy taką oto analogię. Mogę wykluczyć, że jestem w Matriksie, gdyż termin „Matriks” odnosi się do określonego systemu, który widziałem w filmie, w postrzeganym przeze mnie świecie. Nie mogę być jednak akurat w tym systemie, istnieje on bowiem w obrębie postrzeganego przeze mnie świata. A zatem hipoteza „jestem w Matriksie” musi być fałszywa.

Wniosek na temat Matriksa wydaje się słuszny, ale możemy znaleźć na niego dość oczywistą odpowiedź. Powyższy argument wyklucza, że żyję w Matriksie, wciąż jednak dopuszcza możliwość, że przebywam w matriksie, o ile „matriks” rozumiemy jako termin ogólny, określający komputerową symulację świata, w przeciwieństwie do terminu „Matriks”, który oznacza konkretny system, przedstawiony w filmie.

Na tej samej podstawie mogę wykluczyć, że jestem mózgiem w kadzi (jeżeli „mózg” rozumieć jako określony rodzaj systemu biologicznego, występujący w zmysłowo danym świecie). Nie mogę jednak wykluczyć, że jestem zamknięty w kadzi, gdyż oznacza to tylko tyle, że mam system poznawczy, który otrzymuje bodźce i wysyła reakcje do komputerowej symulacji świata. Termin „zamknięty w kadzi” (oraz terminy użyte do jego zde-

finiowania) jest ogólny i niezwiązany z określonymi systemami w danej zmysłowo rzeczywistości. Użycie nieco innego języka pozwala zatem reformułować hipotezę sceptyczną w taki sposób, by była ona odporna na zarzut Putnama.

Mówiąc bardziej technicznie, argument Putnama jest skuteczny, ponieważ „mózg” i „Matriks” to – odpowiednio – naturalny termin rodzajowy oraz nazwa własna. Tego typu terminów dotyczy eksperyment myślowy z bliźniaczą Ziemią (PUTNAM 1975), w którym dwaj bliźniacy nazywają rzeczy tak samo, mimo że inne jest ich odniesienie. Na Ziemi termin „woda” odnosi się do H₂O, ale na bliźniaczej Ziemi (gdzie w oceanach i jeziorach znajduje się z wyglądu ta sama substancja, lecz o wzorze chemicznym XYZ) termin „woda” odnosi się do XYZ. Analogicznie mój termin „mózg” odnosi się do biologicznych mózgów, podczas gdy ten sam termin używany przez istotę zamkniętą w kadzi odnosi się do mózgów wirtualnych. Jeżeli zatem zamknięta w kadzi istota mówi „jestem mózgiem w kadzi”, nie odnosi się ona do biologicznego mózgu, przez co jej twierdzenie jest fałszywe.

Eksperyment z bliźniaczą Ziemią nie dotyczy jednak wszystkich terminów. Wyłamują się z niego zwłaszcza terminy semantycznie neutralne (przynajmniej kiedy używamy ich poza kontekstem semantycznym), takie jak „filozof”, „przyjaciół” i wiele innych. Należą do nich również występujące w tym artykule terminy, takie jak „matriks” czy „zamknięty w kadzi”. Jeżeli przy formułowaniu danej hipotezy stwierdzamy, że „jestem w matriksie”, a nie „jestem w Matriksie”, albo „jestem zamknięty w kadzi” zamiast „jestem mózgiem w kadzi”, rozumowanie Putnama przestaje nam zagrażać. Nawet jeśli mózg w kadzi nie może prawdziwie pomyśleć: „jestem mózgiem w kadzi”, może on prawdziwie myśleć: „jestem w kadzi”. Argument Putnama odwraca tylko naszą uwagę od meritum problemu.

UWAGA 2. Mimo tej różnicy zdań konkluzja mojego tekstu zbieżna jest z innym poglądem Putnama, zgodnie z którym przekonania mózgu w kadzi mogą być prawdziwe, a to dlatego, że odnosi się on do procesów chemicznych albo stanów komputera. Niemniej jednak ja docieram do tego wniosku nieco inną drogą. Putnam odwołuje się do przyczynowej teorii odniesienia: myśli odnoszą się do tego, z czym są przyczynowo powiązane, myśli zaś istoty zamkniętej w kadzi łączą się przyczynowo z procesami zachodzącymi w komputerze. Argument ten jest oczywiście niekonkluzywny, ponieważ przyczynowa teoria odniesienia jest zupełnie nieograniczona. Powiedzieć, że odniesienie wymaga powiązania przyczynowego, to nie to samo, co stwierdzić, jaki jego rodzaj jest wystarczający. W wielu przypadkach (np. „flo-

giston”) terminy posiadają wiele powiązań przyczynowych, choć niczego nie oznaczają. Intuicyjnie traktujemy mózg w kadzi jako przypadek tego rodzaju, stąd odwołanie do przyczynowej teorii odniesienia nie może w niczym pomóc.

Mój argument funkcjonuje niezależnie od teorii odniesienia, gdyż bezpośrednio dotyczy on hipotez pierwszego rzędu na temat świata, połączeń między nimi i konsekwencji, jakie niesie ze sobą ich prawdziwość. Odpowiadając na zarzuty, odwoływałem się do pewnych twierdzeń na temat odniesienia, zgodnych z teorią przyczynową. Co jednak ważne, twierdzenia te były wyłącznie konsekwencjami argumentu pierwszego rzędu, nie zaś jego założeniami. Uważam za ogólną zasadę, że twierdzenia w teorii odniesienia należy dopasowywać do poszczególnych sądów pierwszego rzędu, a nie na odwrót.

UWAGA 3. Sformułowanie „hipoteza sceptyczna” ma w tym artykule znaczenie techniczne. Hipoteza sceptyczna (względem jakiegoś przekonania, że P) to hipoteza, której: (a) nie możemy jednoznacznie wykluczyć; (b) jeżeli ją zaakceptujemy, zmuszeni jesteśmy odrzucić przekonanie, że P. Hipoteza sceptyczna w odniesieniu do jakiejś klasy przekonań to taka hipoteza, która jest sceptyczna względem większości lub wszystkich przekonań tej klasy. Hipoteza globalnie sceptyczna to hipoteza sceptyczna odnosząca się do wszystkich przekonań empirycznych.

Hipoteza sceptyczna (odnośnie do przekonania) zagraża danemu przekonaniu, gdyż nie możemy z pewnością jej wykluczyć, a ponieważ implikuje ona negację owego przekonania, wydaje się (uwzględniając wiarygodną zasadę domknięcia dotyczącą pewności), że nasza wiedza na jego temat nie może być pewna. Jeżeli do tego nie wiemy, czy dana hipoteza sceptyczna nie zachodzi (a tak właśnie jest z większością hipotez przedstawionych w tym artykule), z zasady domknięcia wynika, że przekonania tej klasy nie konstituują wiedzy.

Niekiedy terminu „hipoteza sceptyczna” używa się w szerszym sensie, w odniesieniu do każdej hipotezy, której uwzględnienie sprawia, że nie wiem, czy P. (Hipoteza, w której moje przekonanie jest przypadkowo prawdziwe, jest sceptyczna w tym sensie, w przeciwieństwie do rozumienia podanego wyżej). Nie argumentowałem jednak, że Hipoteza Matriksa nie jest sceptyczna w tym znaczeniu. Chodziło mi raczej o wykazanie, że nawet jeżeli ją uwzględnimy, nasze przekonania są prawdziwe – nie broniłem jednak poglądu, że konstituują one wiedzę. Jestem jednak skłonny, by sądzić, że jeśli mamy wiedzę w normalnym, niebędącym matriksem świecie, możemy mieć również wiedzę w matriksie.

UWAGA 4. Czym jest odpowiednia klasa przekonań? Są oczywiście przekonania – że istnieją, że dwa plus dwa daje cztery albo że nie ma jedno-

rozców – których nie może podważyć nawet hipoteza sceptyczna, negująca istnienie świata zewnętrznego. Z tego względu najlepiej ograniczyć się do przekonań, które: (a) dotyczą świata zewnętrznego; (b) nie są uzasadnialne *a priori*; (c) orzekają coś pozytywnego (nie mogą być prawdziwe w pustym świecie). Dla celów tego artykułu przekonania te możemy nazwać „empirycznymi”. Twierdzenia na temat hipotez sceptycznych, kwestionujących nasze przekonania, powinny być generalnie ograniczone do przekonań tej klasy.

UWAGA 5. Odnośnie do Hipotezy Obliczeniowej: spójne jest przypuszczenie, że fizykę konstituuje poziom obliczeniowy, ale nie jest jasne, czy spójnie można orzec, iż jest on fundamentalny. Jeżeli tak, nasz świat staje się światem „czystych bitów”. Byłby on zatem światem czystych różnic: istniałyby w nim dwa różne, bazowe stany, bez żadnej dalszej różnicy na głębszym poziomie natury. To, czy twierdzenie takie uzna się za spójne, łączy się z odpowiedzią na pytanie, czy każda różnica musi być ugruntowana w jakiejś podstawowej naturze wewnętrznej, czy wszystkie dyspozycje muszą mieć kategorię bazy itd. Ten artykuł nie wymaga jednak poruszania tych kwestii. W Hipotezie Matriksa obliczenie jest implementowane przez procesy zachodzące w świecie stwórcy. Oznacza to, że istnieje pewien bardziej fundamentalny poziom własności wewnętrznych, które służą jako baza dla różnic pomiędzy bitami.

UWAGA 6. Odnośnie do Hipotezy Umysł-Ciało: warto zauważyć, że Hipoteza Matriksa to konkretny przykład tego, w jaki sposób kartezjański dualizm substancjalny mógłby okazać się prawdziwy. Niekiedy uważa się, że pogląd, zgodnie z którym fizyczne procesy oddziałują z нефizycznym umysłem, jest nie tylko mało prawdopodobny, ale wręcz wewnętrznie sprzeczny. Hipoteza Matriksa wprost sugeruje, że jest inaczej. Zgodnie z nią nasz system poznawczy podlega całkiem innym procesom niż te, które zachodzą w fizycznym świecie, choć – z drugiej strony – oddziaływanie między nimi ma charakter przyczynowy.

Koncepcja taka generuje pewne pytania. Na przykład jeżeli zamknięty w kadzi system poznawczy wytwarza ruchowe bodźce cielesne, jaką rolę odgrywa symulowany mózg? Można by się zapewne obyć bez niego, ale powodowałoby to wiele niepożądanych skutków, na przykład w chwili, gdy lekarze w matriksie otwieraliby czaszkę. Rozsądniej przyjąć, że stany mózgu zamkniętego w kadzi i mózgu symulowanego są zawsze izomorficzne – otrzymują one te same bodźce i wytwarzają te same reakcje. Jeżeli oba systemy zaczynają od stanów izomorficznych oraz zawsze otrzymują te same bodźce, wtedy zawsze pozostaną w stanach izomorficznych (indeterminizm

pomijamy). Przy okazji tłumaczy to, dlaczego śmierć w matriksie prowadzi do śmierci w zewnętrznym świecie.

Który z mózgów faktycznie kontroluje ciało? Wszystko zależy od projektu matriksa. Jeżeli reakcji zamkniętego w kadzi systemu nie przekazuje się do symulacji, prawdziwa jest jakaś wersja epifenomenalizmu. Jeżeli impulsy ruchowe symulowanego ciała zależą od reakcji systemu zamkniętego w kadzi, bez uwzględniania reakcji symulowanego mózgu, prawdziwy jest interakcjonizm. Co ciekawe, tego rodzaju stanowisko jest spójne z zasadą przyczynowego domknięcia domeny fizycznej! Zgodnie z trzecią możliwością mechanizmy matriksa uwzględniają reakcje obu układów (być może jakoś je uśredniają). Mogłoby to prowadzić do pewnej zbędności przyczynowania. Możliwe, że układ kontrolny matriksa przełączałby nawet czasem pomiędzy nimi. W każdym z tych przypadków, jeżeli tylko dwa systemy pozostają w stanach izomorficznych, ich zachowanie będzie takie samo.

Można się zastanawiać, czy w takim przypadku nie istniałyby dwa świadome umysły, co przypominałoby historyjkę Daniela Dennetta, znaną jako „Gdzie jestem?”. Odpowiedź zależy od tego, czy do wytworzenia umysłu wystarczą procesy obliczeniowe. Jeśli rację mają przeciwnicy komputacjonizmu (tacy jak John Searle), istnieć będzie jeden umysł. Jeśli jednak rację mają jego zwolennicy, istnieć mogą dwa zsynchronizowane umysły (co z kolei prowadzi do pytania, który z nich jest mój). Pogląd, że posiadamy tylko jeden umysł, jest zdecydowanie bardziej zdroworozsądkowy, ale nie można wykluczyć koncepcji dwóch umysłów.

Jedną z zalet przyjęcia komputacjonizmu jest możliwość rozwinięcia hipotezy, że żyjemy w komputerowej symulacji, mimo że nie posiadamy oddzielnego, zewnętrznie podłączonego układu poznawczego. Twórcy uruchamialiby po prostu symulację, obejmującą również symulację mózgów, na bazie których powstawałyby umysły. To zdecydowanie ułatwiłoby im pracę, gdyż usuwałoby konieczność tworzenia i podtrzymywania systemów poznawczych. Z tego względu w przyszłości najprawdopodobniej istnieć będzie wiele symulacji tego rodzaju, nie jest zaś oczywiste, czy powstaną symulacje tak nieporęczne, jak ta przedstawiona w *Matriksie*. (Dlatego argument Bostroma, że możemy żyć w symulacji stosuje się raczej do „oszczędniejszej” jej wersji, a nie do tej, którą zaprezentowano w filmie). Taka hipoteza odpowiada Hipotezie Metafizycznej, z której usunięto Hipotezę Umysł-Ciało. Podobnie jak w innych przypadkach nie jest to hipoteza sceptyczna: jeśli jesteśmy w takiej symulacji (i jeśli komputacjonizm to poprawny pogląd na naturę umysłu), większość naszych przekonań o zewnętrznym świecie wciąż jest poprawna.

Są też inne możliwości. Jedna z bardziej intrygujących (omawia ją CHALMERS 1990) ma swe źródło w badaniach nad sztucznym życiem, w których wykorzystuje się stosunkowo prostą symulację środowiska oraz złożone reguły jego interakcji z symulowanymi stworzeniami. Algorytmy odpowiedzialne za procesy „umysłowe” stworzeń są całkowicie niezależne od tych, które zarządzają „fizyką” środowiska. Mieszkańcy takiej symulacji prawdopodobnie nigdy nie znaleźliby w swoim świecie źródła własnych procesów poznawczych. Jeżeli staliby się oni naukowcami, przyjęliby kartezjański dualizm, utrzymując (poprawnie!), że ich procesy poznawcze mają swe źródło poza światem fizycznym. Oto kolejny sposób, w jaki dualizm może okazać się prawdziwy.

UWAGA 7. Argumentowałem, że Hipoteza Matriksa implikuje Hipotezę Metafizyczną i na odwrót. Implikacja jest w tym znaczeniu relacją epistemiczną: jeżeli ktoś akceptuje pierwszą hipotezę, powinien przyjąć również i drugą. Nie twierdzę, że z Hipotezy Metafizycznej logicznie wynika Hipoteza Matriksa w tym sensie, że w każdym świecie możliwym, w którym obowiązuje Hipoteza Matriksa, prawdziwa jest również Hipoteza Metafizyczna. Twierdzenie takie wydaje się fałszywe. Istnieją na przykład światy możliwe, w których fizyczna czasoprzestrzeń nie została przez nikogo stworzona (a zatem Hipoteza Metafizyczna jest fałszywa), ale w których jestem podłączony do sztucznie zaprojektowanej, komputerowej symulacji, zlokalizowanej w obrębie fizycznej czasoprzestrzeni (a zatem poprawna jest Hipoteza Matriksa). Jeżeli fizyka w aktualnym świecie nie jest obliczeniowa, wtedy fizyka w tymże świecie możliwym również nie jest obliczeniowa. Można zatem powiedzieć, że obie hipotezy są równoważne *a priori*, ale nie w sposób konieczny.

(Termin „fizyka”, którego używa moja zamknięta w kadzi jaźń, w innym świecie możliwym odnosi się oczywiście do czegoś, co jest zarazem obliczeniowe i stworzone. Ale termin „fizyka”, którego używam jako jaźń aktualnie zamknięta w kadzi, odnosi się do obowiązującej w świecie zewnętrznym fizyki nieobliczeniowej).

Różnica ta ma swoje źródło w dwóch sposobach traktowania Hipotezy Matriksa: jako dotyczącej tego, co aktualnie możliwe oraz jako dotyczącej tego, co mogłoby zachodzić, ale nie zachodzi. Pierwsza opcja wyrażana jest w oznajmujących okresach warunkowych: jeżeli jestem teraz w matriksie, to mam dłonie, atomy zrobione są z bitów, a Hipoteza Metafizyczna jest prawdziwa. Druga opcja wyrażana jest za pomocą przypuszczających okresów warunkowych: jeśli byłbym w matriksie, to nie miałbym dłoni, atomy nie byłyby zrobione z bitów, a Hipoteza Metafizyczna nie byłaby prawdziwa.

Dostrzec można tu podobieństwo do różnych sposobów ujmowania Putnamowskiego scenariusza z bliźniaczą Ziemią, który często pojawia się w kontekście dyskusji nad semantyką dwuwymiarową. Jeżeli aktualnie jestem w świecie XYZ, wtedy XYZ to woda; jeżeli jednak byłbym w świecie XYZ, XYZ nie byłoby wodą (woda wciąż miałaby wzór H_2O). W pierwszym przypadku rozważamy bliźniaczą Ziemię jako świat aktualny, a w drugim – jako możliwy. Możemy powiedzieć, że bliźniacza Ziemia weryfikuje zdanie „woda to XYZ”, ale spełnia zdanie „woda nie jest XYZ”, przy czym weryfikacja i spełnianie odpowiadają ujmowaniu bliźniaczej Ziemi jako aktualnej i możliwej.

Świat matriksa weryfikuje Hipotezę Metafizyczną, ale nie spełnia jej, bowiem hipoteza ta mówi coś o fizyce i świecie fizycznym, ale to, czym jest „fizyka” różni się w zależności od tego, czy świat matriksa rozumiemy jako aktualny, czy jako możliwy. Jeżeli jestem w matriksie, fizyka jest obliczeniowa. Ale jeżeli byłbym w matriksie, fizyka nie byłaby obliczeniowa (obliczeniowy byłby matriks, w przeciwieństwie do komputera i mojego mózgu, które opierałyby się na fizyce niezależnej od procesów obliczeniowych). W ten sposób twierdzenia na temat fizyki i procesów fizycznych, wygłaszane w świecie matriksa, są analogiczne do twierdzeń na temat wody, wygłaszanych na bliźniaczej Ziemi.

UWAGA 8. Odpowiedzi na pierwsze zarzuty, wyszczególnione w punkcie 7 są w sposób oczywisty związane z przyczynowym ujęciem odniesienia. Zgodnie z tym, co powiedziałem, prawdziwość myśli zamkniętej w kadzi istoty nie zależy od jej bezpośredniego środowiska, ale od tego, z czym ta istota jest przyczynowo powiązana, tzn. od komputerowych procesów, do których jest podłączona. Jak zauważyłem wcześniej, do takiego wniosku można dojść, wykorzystując argument pierwszego rzędu, bez zakładania przyczynowej teorii odniesienia. Kiedy jednak osiągnie się tę konkluzję, ujawniają się różne ciekawe związki.

Na przykład mój termin „włosy” odnosi się do włosów, mój zaś zamknięty w kadzi odpowiednik odnosi się za jego pomocą do wirtualnych włosów, przy czym oba twierdzenia mają podobną strukturę. Zachodzi tu strukturalna analogia z bliźniaczą Ziemią, na której Oskar (mieszkaniec Ziemi) odnosi się do wody (H_2O), podczas gdy bliźniak Oskara (mieszkaniec bliźniaczej Ziemi) odnosi się do bliźniaczej wody (XYZ). W obu przypadkach terminy odnoszą się do tego, z czym przyczynowo się łączą. Te naturalne terminy rodzajowe funkcjonują przez oznaczanie konkretnych fragmentów środo-

wiska podmiotu, dokładna zaś natura tych rodzajów zależy od natury środowiska. Coś podobnego zastosować można do nazw własnych, np. „Tucson”.

Zachowanie tych terminów modelują dwuwymiarowe ramy semantyczne. Kiedy ujmujemy bliźniaczą Ziemię jako świat aktualny, weryfikuje on zdanie „woda to XYZ”, gdy zaś rozważamy ten świat jako możliwy, spełnia on zdanie „woda nie jest XYZ”. Analogicznie, kiedy świat matriksa traktujemy jako aktualny, weryfikuje on zdanie „włosy zrobione są z bitów”, kiedy zaś traktujemy go jako tylko możliwy, spełnia on zdanie „włosy nie są zrobione z bitów”.

Różnica między ujmowaniem świata jako aktualnego lub możliwego powoduje zmianę perspektywy, którą wskazywałem w odpowiedzi na zarzut 5. Jeśli świat matriksa traktowany jest jako czysto możliwy, poprawne jest stwierdzenie, że jego mieszkańcy nie mają włosów (mają wyłącznie wirtualne włosy). Kiedy jednak świat matriksa rozważamy jako aktualny (tzn. jeśli hipotetycznie przyjmujemy, że jesteśmy w matriksie), powinniśmy stwierdzić, że istoty go zamieszkujące mają włosy, jakkolwiek wirtualne.

Analogia z bliźniaczą Ziemią może sugerować, że znaczenie terminów, takich jak „włosy” oraz treść odpowiadających im myśli zależy od naszego środowiska. Ale podejście dwuwymiarowe sugeruje ponadto, że istnieje pewien, niezależny od otoczenia, wewnętrzny aspekt treści, wspólny dla obu bliźniaków. Intensja pierwotna zdania jest prawdziwa w jakimś świecie, gdy świat ten weryfikuje to zdanie, intensja wtórna jest zaś prawdziwa w jakimś świecie, gdy świat ten spełnia to zdanie. Zdanie zatem wypowiediane przez Oskara i jego bliźniaka – „woda jest mokra” – ma różne intensje wtórne (które, mówiąc z grubsza, są prawdziwe, kiedy mokre jest zarówno H₂O, jak i XYZ), ale te same intensje pierwotne (prawdziwe w światach, gdzie to, co wodniste, jest mokre). Podobnie zdanie „mam włosy”, używane przeze mnie i mój zamknięty w kadzi odpowiednik, ma różne intensje wtórne (prawdziwe w światach, w których mamy biologiczne lub komputerowe włosy), ale te same intensje pierwotne (prawdziwe w światach, gdzie istnieje coś, co wygląda jak włosy). Pierwotne intensje naszych myśli i języka reprezentują istotny, wspólny wymiar ich treści.

UWAGA 9. Dlaczego jednak, jak wskazywałem w odpowiedzi na zarzut 7, inaczej jest z terminami takimi jak „działanie” czy „przyjaciół”? Już wcześniej (w uwadze 1) zauważyliśmy, że nie wszystkie terminy funkcjonują tak jak „woda” czy „włosy”. Istnieje pewna grupa terminów semantycznie neutralnych, które nie podpadają pod eksperyment myślowy z bliźniaczą Ziemią: każdy z bliźniaków, choć żyje w innym środowisku, będzie używał

tych terminów w tym samym znaczeniu (o ile używa ich poza kontekstem semantycznym). Terminy te to na przykład „oraz”, „przyjaciół”, „filozof”, „działanie”, „doświadczenie” czy „zamknięty w kadzi”. Choć zatem terminy istoty zamkniętej w kadzi, takie jak „ręka”, „włosy” czy „Tucson”, mogą znaczyć coś innego niż nasze terminy, „przyjaciół”, „filozof” czy „działanie”, zachowają dla nas takie samo znaczenie.

Wynika z tego, że jeśli rozważamy przekonania zamkniętej w kadzi istoty takie jak „mam przyjaciół” czy „wykonuję działania”, nie możemy wykorzystać odpowiedzi, jaką sugeruje eksperyment z bliźniaczą Ziemią. Przekonania te będą prawdziwe wtedy i tylko wtedy, gdy istota ta ma przyjaciół i wykonuje jakieś działania. Na szczęście można rozsądnie twierdzić, że *na prawdę* ma ona przyjaciół i *na prawdę* wykonuje działania (w swoim środowisku, nie w naszym). To samo dotyczy innych terminów semantycznie neutralnych: to właśnie ta klasa wyrażań, dla których powyższa odpowiedź jest zasadna.

UWAGA 10. Jaka jest ontologia przedmiotów wirtualnych? To trudne pytanie, choć nie bardziej niż pytanie o ontologię zwyczajnych przedmiotów makroskopowych w świecie, którym rządzą prawa mechaniki kwantowej. Odpowiedź na zarzut 6 sugeruje, że w obu przypadkach powinniśmy odrzucić twierdzenia o egzemplarycznej identyczności pomiędzy poziomem makroskopowym i mikroskopowym. Stoły nie są identyczne z żadnym przedmiotem, charakteryzowanym w języku mechaniki kwantowej, podobnie jak wirtualne stoły nie są identyczne z przedmiotami charakteryzowanymi jako bity. Niemniej fakty o stołach superwenują na faktach poziomu mechaniki kwantowej, a fakty o wirtualnych stołach superwenują na faktach poziomu obliczeniowego. Stoły są konstytuowane przez procesy kwantowe, a wirtualne stoły – przez procesy obliczeniowe. Dalsze uszczegółowienie tych przypadków zależy od delikatnych kwestii metafizycznych.

Biorąc pod uwagę ujęcie trzecioosobowe, w którym spoglądamy na mózg w kadzi z perspektywy naszego świata, można by stwierdzić, że przedmioty wirtualne tak *naprawdę* nie istnieją: nie są to prawdziwe przedmioty odpowiadające czemukolwiek, co istnieje w komputerze. Takie twierdzenie może jednak zmuszać do przyjęcia poglądu, że w naszym świecie, w którym obowiązuje mechanika kwantowa, przedmioty te również nie istnieją. Jeżeli ktoś stosuje ograniczoną ontologię przedmiotów w jednym przypadku, powinien zastosować ją również w drugim; podobnie rzecz się ma ze stosowaniem ontologii bardziej liberalnych. Przypadki te można potraktować oddzielnie tylko wtedy, kiedy przyjmie się jakąś wersję kontekstualizmu,

w którym to, czy coś podpada pod kategorię „przedmiotu” (albo domenę kwantyfikatora takiego jak „wszystko”), uzależnione jest od kontekstu, który ma na myśli mówiący. Jest to jednak tylko odzwierciedlenie mało ważnego faktu o naszym języku, nie zaś jakiejś głębokiej prawdy o świecie. W istotnych względach przedmioty wirtualne są nie mniej rzeczywiste niż ich zwyczajne odpowiedniki.

UWAGA 11. Odpowiedź na zarzut 8 dotyka kwestii, związanej z poglądami Russella i Kanta, którzy utrzymywali, że nie znamy wewnętrznej natury bytów, istniejących w zewnętrznym świecie. Jeżeli chodzi o własności fizyczne, percepcja i nauka mogą powiedzieć nam, jak te byty oddziałują na nas i w jakiej relacji pozostają do siebie, nie mówią nam jednak, jakie są one w sobie. Innymi słowy, metody te informują o przyczynowej strukturze świata zewnętrznego, milczą natomiast jeżeli chodzi o naturę tego, co wewnętrzne.

Hipoteza Metafizyczna częściowo mówi o tym, co jest podstawą tej mikrofizycznej struktury przyczynowej: byty mikrofizyczne zbudowane z bitów. To samo dotyczy Hipotezy Matriksa. Można zatem stwierdzić, że jeżeli jesteśmy w matriksie, kantowska *Ding-an-sich* (rzecz w sobie) jest częścią komputera-w-sobie! Ta hipoteza uzupełnia nasze potoczne rozumienie świata, a zarazem nie jest z nim sprzeczna, gdyż pojęcia te nie informują o wewnętrznej naturze rzeczywistości.

UWAGA 12. Ogólny morał jest taki: nasz „potoczny obraz świata” trzyma się mocno – zdroworozsądkowa koncepcja poziomu makroskopowego nie daje się łatwo sfalsyfikować przez odkrycia naukowe i metafizykę. Jeżeli tylko świat fizyczny obejmuje procesy o właściwej strukturze przyczynowej i kontrfaktycznej, będzie on zgodny z naszym oczywistym jego obrazem. Nawet komputerowa symulacja ma odpowiednią strukturę przyczynową i kontrfaktyczną, podobnie jak procesy w umyśle Boga: dlatego dostarczają one argumentów na rzecz istnienia zewnętrznej rzeczywistości, mimo że jej natura jest zaskakująca.

Tego rodzaju zmienność w naszym rozumieniu świata jest blisko związana z semantycznym brakiem neutralności wielu naszych pojęć. Niektóre z nich, takie jak „woda”, „włosy” czy „elektron”, pozwalają na zmianę odniesienia, w zależności od tego, czym się ono okaże. Odniesienia te pojmujemy wyłącznie poprzez aktualnie odgrywaną rolę przyczynową albo wygląd, ale bez odpowiedzi pozostawiamy kwestię ich wewnętrznej natury. Można argumentować, że największe ograniczenia nakładane przez nasze rozumienie świata dotyczą pojęć semantycznie neutralnych, które nie pozwalają na tego rodzaju zmienność. Obejmują one wiele naszych pojęć przy-

czynowych (i nomicznych) oraz znaczną część pojęć odnoszących się do umysłu. W tych przypadkach dysponujemy „bezpośrednim” rozumieniem tego, jaki musi być świat, by spełniał te pojęcia. Jeżeli tak, to nasze przekonania co do przyczynowości i umysłu nakładają mocne ograniczenia na to, jaki musi być aktualny świat.

Podstawowe pojęcia semantycznie neutralne mają charakter umysłowy („doświadczenie”, „przekonanie”), przyczynowy („przyczyna”, „prawo”), logiczny i matematyczny („i”, „dwa”) czy kategoriowy („przedmiot”, „własność”). Istnieją również pojęcia neutralne semantycznie, które należą do kilku grup, np. „przyjaciel”, „działanie” czy „komputer”. Fundamentalne ograniczenia, jakie nasze przekonania nakładają na świat zewnętrzny, polegają zatem na tym, że świat ten obejmuje odpowiednie stany mentalne (nasze i cudze) oraz przedmioty i własności, które w stosunku do siebie oraz do stanów mentalnych pozostają w odpowiednich relacjach przyczynowych. Taka koncepcja jest wystarczająco słaba, by mogła zostać spełniona przez matriks (o ile podłączonych jest do niego wiele kadzi albo komputacjonizm jest poprawną teorią umysłu).

Uważam, że kwestia fundamentalnych ograniczeń, które nasze przekonania nakładają na świat, jest najgłębszym zagadnieniem filozoficznym, powstającym przy okazji rozważań nad matriksem. Jeżeli to, co dotąd powiedziałem, jest słuszne, to dzieje się tak dlatego, że ograniczenia te są na tyle słabe, iż wiele hipotez, które można by potraktować jako „sceptyczne”, okazuje się spójnych z naszymi przekonaniem. To zaś pozwala wysunąć pewnego rodzaju odpowiedź na wyzwanie, rzucone przez sceptyków. Nieco paradoksalnie można by stwierdzić, że odpowiedź taka jest możliwa, gdyż wymagamy tak niewiele, by wiedzieć tak dużo.

UWAGA 13. Dlaczego komputerowa symulacja świata spełnia te wymogi? Odpowiedź związana jest z naturą obliczenia i implementacji. Dowolne obliczenie formalne można potraktować jako podstawę specyfikacji (abstrakcyjnej) struktury przyczynowej, ściśle określającej sposób oddziaływania pomiędzy pewnymi zbiorami stanów formalnych. Implementacja takiego formalnego obliczenia wymaga tego, by miała ona konkretne stany, bezpośrednio odzwierciedlające te stany formalne, gdzie wzór interakcji (przyczynowej lub kontrfaktycznej) pomiędzy owymi stanami dokładnie odwzorowuje wzór interakcji pomiędzy stanami formalnymi (zob. CHALMERS 2011). A zatem dowolne dwie implementacje danego obliczenia dzielą określoną strukturę przyczynową. Od obliczeniowego opisu fizycznego świata wymaga się, by odwzorowywał jego strukturę przyczynową aż do poziomu fundamen-

talnych przedmiotów i własności. Każda implementacja tego obliczenia będzie zatem przejawiała tę strukturę przyczynową (w przejściach między stanami implementowanymi, bez względu na to, czy będą to wolty, obwody, czy coś zupełnie innego). Jeżeli więc nasza koncepcja zewnętrznego świata nakłada ograniczenia na strukturę przyczynową, które rzeczywisty świat fizyczny może spełnić, będą one również spełniane przez komputerową symulację.

(Kwestia ta związana jest z uwagą Huberta Dreyfusa, która pojawia się w artykule z tego zbioru⁹. Podobnie jak ja, Dreyfus przyjmuje, że większość przekonań mieszkańców matriksa jest prawdziwa. Sugeruje on jednak, że fałszywych jest wiele ich przekonań przyczynowych, np. ogólne przekonanie o „fizycznym wszechświecie z mocami przyczynowymi, które odpowiadają za zdarzenia w naszym świecie” i pewne węższe przekonania: że bakterie powodują choroby, że słońce powoduje, iż coś się nagrzewa itd. W moim ujęciu sugestia ta jest błędna. Twierdzę bowiem, że w matriksie przyczynowanie, będąc ugruntowane w realnym przyczynowaniu wewnątrz komputera, naprawdę zachodzi; wirtualne bakterie naprawdę powodują wirtualne choroby. Jeżeli zatem mieszkańcy matriksa stwierdzają: „bakterie powodują choroby”, mają rację).

Oczywiście spełnione muszą również zostać ograniczenia dotyczące pojęć umysłowych. Szczególnie ważne jest, by struktura przyczynowa znajdowała się we właściwej relacji do naszego doświadczenia. Ale ograniczenia te będą spełnione również pod warunkiem, że żyjemy w wielokadziowym matriksie lub jeśli komputacjonizm to prawdziwa teoria umysłu. W ten sposób matriks ma wszystko, co potrzeba, by spełnić kluczowe ograniczenia przyczynowe i umysłowe, występujące w naszych pojęciach na temat świata.

UWAGA 14. Możliwa linia zarzutów pod adresem zaproponowanego w tym artykule argumentu dotyczy istnienia dalszych ograniczeń, które nasze przekonania nakładają na świat, a których Hipoteza Matriksa nie spełnia. Można na przykład utrzymywać, że samo dopasowanie własności umysłowych i przyczynowych nie wystarczy, gdyż świat musi mieć odpowiednie własności przestrzenne, które możemy uchwycić bezpośrednio (być może dlatego, że pojęcia przestrzenne mają charakter semantycznie neutralny). Problem w tym, że własności przestrzenne to fikcja: sądzimy, że byty zewnętrzne tworzą pewien wzór przestrzenny, mimo że we wnętrzu komputera takiego wzoru nie ma.

⁹ Zob. Hubert DREYFUS, „Existential Phenomenology and the Brave New World of The Matrix”, *The Harvard Review of Philosophy* 11 (2003), 1: 18–31. (Przyp. tłum.).

W odpowiedzi można stwierdzić, że tego typu dalsze ograniczenia po prostu nie istnieją. Ponieważ pojęcia przestrzenne są przedmiotem eksperymentu myślowego z bliźniaczą Ziemią, nie są one semantycznie neutralne. Mój student Brad Thompson rozwinął eksperyment myślowy tego typu (THOMPSON 2003), oparty na pojęciach podwójnej Ziemi, na której „jeden metr” odnosi się do tego, co my nazywamy dwoma metrami, czy świata El Greco, gdzie „kwadrat” odnosi się do tego, co nazywamy trójkątami. W takim wypadku nasze pojęcia przestrzenne oznaczają dowolne własności i relacje, które w zewnętrznym świecie odpowiadają przyczynowo za związane z nimi doświadczenia przestrzeni: w tym względzie pojęcia przestrzenne są analogiczne do pojęć kolorów. Nie dysponujemy „bezpośrednim” ujęciem podstawowej natury własności przestrzennych. Ponownie okazuje się, że fundamentalne ograniczenia mają charakter przyczynowy i umysłowy.

Zarzut tego rodzaju pojawia się między wierszami w paragrafie 8, gdzie sugeruję, że jeżeli poziom obliczeniowy poniżej poziomu fizyki naprawdę istnieje, to każda implementacja odpowiedniego obliczenia formalnego może w zasadzie służyć jako realizacja tego poziomu, bez naruszania rzeczywistości fizycznej. Oponent mógłby zaprzeczyć, że poziom obliczeniowy istnieje, lub przynajmniej wskazać ograniczenia co do rodzaju implementacji, jaki może okazać się użyteczny. Można na przykład twierdzić, że poziom obliczeniowy sam musi mieć odpowiednią strukturę przestrzenną.

Wydaje mi się, że taka odpowiedź sprzeciwia się duchowi współczesnej fizyki. Fizycy poważnie traktują ideę, że to, co rozumiemy jako przestrzeń, nie jest własnością fundamentalną i że leżącego niżej poziomu nie da się opisać przy użyciu pojęć przestrzennych. Hipoteza komórkowej automatyzacji to tylko jedna z propozycji. Kluczową rolę odgrywa wzór oddziaływania przyczynowego. Jeśli fizycy odkryją, że wzór ten jest realizowany przez zupełnie inny poziom z całkowicie różnymi własnościami, nie wyciągną oni wniosku, że przestrzeń fizyczna nie istnieje. Będą oni raczej wnosić, że przestrzeń jest konstituowana przez coś, co samo nie jest przestrzenne. Tego rodzaju odkrycie może być zaskakujące i prowadzić do korekt w obrębie fizyki, ale nie w większym stopniu niż mechanika kwantowa. Podobnie jak w przypadku tej ostatniej prawie na pewno nie rozumiano by tej hipotezy jako sceptycznej w odniesieniu do makroskopowego świata zewnętrznego. Jeżeli mam rację, nasze pojęcia odnoszące się do świata makroskopowego nie nakładają żadnych istotnych ograniczeń na podstawowy poziom świata, jeżeli chodzi o przestrzeń.

Podobne kwestie pojawiają się w przypadku czasu, który generuje jednak mniej problemów, ponieważ komputerowa symulacja w matriksie rozwija się w porządku czasowym, i to takim samym jak w symulowanym świecie. Nie można zatem stwierdzić, jak w przypadku przestrzeni, że struktura czasowa nie występuje w matriksie. Nawet jeśli pojęcia czasowe byłyby semantycznie neutralne, Hipoteza Matriksa podtrzymywałaby nasze przekonania na temat czasu. Wciąż jednak można wskazywać, iż nasze pojęcia odnoszące się do zewnętrznego czasu nie są semantycznie neutralne (warto odnotować, że fizycy rozwijają hipotezy, w których pojęcia czasowe nie odgrywają żadnej roli na poziomie fundamentalnym). Oznaczają one raczej zewnętrzne bogactwo własności i relacji odpowiadających za analogiczną różnorodność naszych doświadczeń czasowych. Jeżeli tak, każda symulacja komputerowa, posiadająca poprawną strukturę przyczynową i właściwą relację do naszego doświadczenia, uzasadniać będzie przekonania na temat czasu, bez względu na jego wewnętrzną naturę.

UWAGA 15. Przedstawione w tym artykule rozumowanie nie dostarcza jednoznacznego obalenia sceptycyzmu; kilka odmian tego stanowiska wciąż jest dostępnych. Myślę jednak, że znacznie wzmacnia ono standardowe odpowiedzi na sceptycyzm. Często sądzi się, że choć różne hipotezy sceptyczne można pogodzić z naszym doświadczeniem, hipoteza o istnieniu realnego świata fizycznego zapewnia prostsze czy lepsze wyjaśnienie regularności naszego doświadczenia. A jeżeli tak, to dzięki odwołaniu się do najlepszego wyjaśnienia, uzasadniona jest wiara w istnienie realnego świata fizycznego.

W tym miejscu często pojawia się zarzut, że niektóre hipotezy sceptyczne – głoszące na przykład, że całe nasze doświadczenie ma swe źródło w komputerowej symulacji albo w Bogu – wydają się równie proste, co standardowe wyjaśnienie, przez co zaproponowana odpowiedź na sceptycyzm zawodzi. Jeżeli jednak mam rację, te „równie proste” hipotezy nie są w ogóle sceptyczne. Wciąż działa zatem rozumowanie do najlepszego wyjaśnienia: wszystkie te „proste” hipotezy głoszą, że większość twierdzeń na temat zewnętrznego świata jest prawdziwa.

Ostatnia kwestia dotyczy hipotez sceptycznych, takich jak Hipoteza Niedawnego czy Lokalnego Matriksa, które wciąż można brać pod uwagę. Wydają się one jednak znacznie bardziej skomplikowane niż hipotezy wspomniane wyżej, gdyż proponują niejednorodne wyjaśnienie regularności naszego doświadczenia. Hipoteza Niedawnego Matriksa głosi, że obecne i przeszłe regularności mają całkowicie różne wyjaśnienie. Hipoteza Lokalnego Matriksa broni tego samego twierdzenia w odniesieniu do zdarzeń blis-

kich i oddalonych. Wzięte razem, hipotezy te bazują na mechanizmach dualistycznych i są dalece bardziej złożone niż jednorodne propozycje, zarysowane powyżej. Wnioskowanie do najlepszego wyjaśnienia pozwala zatem na wykluczenie tych hipotez i przyjęcie ich niesceptycznych wersji.

Nawet jeżeli uważamy, że hipotezy sceptyczne dostarczają całkiem dobrego wyjaśnienia naszego doświadczenia, wciąż dysponujemy obiecującym argumentem przeciwko globalnej hipotezie sceptycznej, kwestionującej istnienie świata zewnętrznego. Jeżeli mam rację, wszystkie te hipotezy mają w najgorszym razie charakter częściowo sceptyczny: ich poprawność nie kwestionuje znacznej części przekonań empirycznych ani istnienia zewnętrznego świata. Aby móc przejść do globalnego sceptycyzmu, należy przyjąć Hipotezę Chaosu, w której regularności naszego doświadczenia w ogóle nie mają wyjaśnienia. Nawet najśłabsza wersja wniosku do najlepszego wyjaśnienia eliminuje jednak pogląd tego rodzaju. Jeżeli tak, zasada ekonomii myślenia może uzasadniać nasze przekonanie co do istnienia świata zewnętrznego.

Z języka angielskiego przełożył Jacek Jarocki

BIBLIOGRAFIA

- BOSTROM, Nick. 2003. „Are you living in a computer simulation?”. *Philosophical Quarterly* 53 (211): 243–255.
- CHALMERS, David. 1990. „How Cartesian dualism might have been true”, niepublikowany.
- CHALMERS, David. 2011. „A computational foundation for the study of cognition”. *Journal of Cognitive Science* 12 (4): 323–357.
- DENNETT, Daniel. 1978. „Where am I?”. W: Daniel DENNETT. *Brainstorms*, 310–323. (Boston, Mass.: MIT Press. (Autoryzowany, polski przekład, zatytułowany *Gdzie jestem?*, dostępny jest online: http://autofocus.blox.pl/resource/gdzie_jestem.pdf).
- PUTNAM, Hilary. 1975. „The meaning of ‘meaning’”. W: Hilary PUTNAM. *Mind, Language, and Reality*. Cambridge: Cambridge University Press. (Wydanie polskie: Hilary PUTNAM. „Znaczenie wyrazu ‘znaczenie’”. W: Hilary PUTNAM. *Wiele twarzy realizmu i inne eseje*, przełożył Adam Grobler, 93–184. Warszawa: PWN, 2014).
- PUTNAM, Hilary. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- SEARLE, John. 1984. „Can computers think?”. W: John SEARLE. *Minds, Brains, and Science*. Harvard: Harvard University Press, 28–41. (Wydanie polskie: John SEARLE. „Czy komputery mogą myśleć?”. W: John SEARLE. *Umysł, mózg i nauka*, przełożył Jerzy Bobryk. Warszawa: PWN, 1995, 25–37).

Mgr JACEK JAROCKI – doktorant w Katedrze Historii Filozofii Nowożytnej i Współczesnej na Wydziale Filozofii KUL; adres do korespondencji: Al. Raławickie 14, 20-950 Lublin; e-mail: jacekjarocki@kul.pl

THOMPSON, Bradley Jon. 2003. *The Nature of Phenomenal Content*, niepublikowana rozprawa doktorska, University of Arizona.

WOLFRAM, Stephen. 2002. *A New Kind of Science*. Champaign: Wolfram Media.

Słowa kluczowe: matriks, metafizyka, sceptycyzm, mózg w kadzi.

Key words: matrix, metaphysics, skepticism, brain in a vat.

Information about Author: Prof. DAVID CHALMERS — Professor of Philosophy and co-director of the Center for Mind, Brain, and Consciousness at the New York University, and also (20% time) Professor of Philosophy at the Australian National University; address for correspondence: Department of Philosophy, 5 Washington Place, New York, NY 10003; e-mail: chalmers@nyu.edu

Information about Translator: JACEK JAROCKI, MA — PhD student, Department of the History of Modern and Contemporary Philosophy, Faculty of Philosophy at the John Paul II Catholic University of Lublin; address for correspondence: Al. Raławickie 14, PL 20-950 Lublin; e-mail: jacekjarocki@kul.pl