Filozofia Nauki
Rok XVII, 2009, Nr 3(67)

Wojciech P. Grygiel
Mateusz Hohol

Rogera Penrose'a kwantowanie umysłu

WPROWADZENIE

Ekspansja współczesnej metody naukowej, opartej na matematycznym modelowaniu oraz eksperymentalnej weryfikacji badanych zjawisk, sięga dziś obszarów rzeczywistości, które jeszcze do niedawna były przedmiotem mglistych spekulacji i czysto zdroworozsądkowych dociekań. O ile siła tej metody uwidoczniła się już szeroko w badaniach nad strukturą makro- i mikroświata, o tyle zagadnienie funkcjonowania ludzkiego mózgu i świadomości pozostaje nadal raczej w kręgu rozbieżnych hipotez niż dobrze uzasadnionych modeli. Jeden z fundamentalnych aspektów tego zagadnienia — problem relacji między umysłem a ciałem (*mind-body problem*) — nurtował filozofów od czasów starożytności i pozostaje nierozwiązany do dziś.

Najstarszym stanowiskiem filozoficznym w tej kwestii jest dualizm substancji, który zakłada istnienie materialnego ciała i niematerialnego umysłu. Przedstawicielami takiego poglądu byli Platon (dusza i ciało)[1] i Kartezjusz (*res extensa* i *res cogitans*).[2] Inną wersją tego stanowiska jest dualizm własności, w którym własności mentalne i fizykalne są różnymi aspektami tej samej substancji. Koncepcje dualistyczne spotyka się w filozofii nawet współcześnie, a do ich głównych przedstawicieli należą Karl Popper i John Eccles.[3] Przenieśli oni dwoistość substancji na poziom neuronowy. Inną próbą rozwiązania problemu *mind-body* jest cała gama stanowisk *materialistycznych*. Najbardziej radykalną odmianą jest *materializm elimina-*

[1] Zob. np. Platon, *Fedon*, [w:] *Dialogi*, przeł. W. Witwicki, Warszawa 1988, PIW.
[2] Kartezjusz, *Medytacje o filozofii pierwszej*, przeł. J. Hartman, Kraków 2006, Zielona Sowa.
[3] K. Popper, J. Eccles, *The self and its brain*, Berlin 1977, Springer.

cyjny, reprezentowany przez Paula Churchlanda.[4] Twierdzi on, iż w rzeczywistości umysł nie istnieje, co implikuje nieprawdziwość psychologii potocznej. Musi ona zostać wyeliminowana na rzecz neuronauki. Odmianą materializmu jest również *behawioryzm analityczny*, który zakłada, że zjawiska mentalne można zredukować do zachowań. Przedstawicielem takiego poglądu był brytyjski filozof Gilbert Ryle.[5] W panoramie stanowisk materialistycznych istotną rolę odgrywają również teorie identyczności. Jednym z ujęć jest koncepcja Johna Jamiesona Carswella Smarta,[6] w której głosi on, iż stany mentalne są tożsame ze stanami mózgu. Przykładowo uczucie zimna w tym wypadku oznacza reakcję odpowiednich części układu nerwowego na zewnętrzne bodźce.

Innym popularnym stanowiskiem jest *funkcjonalizm*, którego twórcą jest Hilary Putnam.[7] Stanowisko to zakłada, że stany mentalne mogą być realizowane przez odpowiednio skomplikowane struktury fizyczne. Współcześnie funkcjonalizm przyjmuje wiele form. Jedną z nich jest silna wersja sztucznej inteligencji (*strong artificial ingelligence*), według której umysł jest formalnym programem czy wręcz systemem operacyjnym „zainstalowanym" w mózgu. Już w 1943 roku Warren McCulloch i Walter Pitts zaproponowali, aby neurony potraktować jako „logiczne narzędzia", mogące przyjmować jedynie dwa stany 0 i 1, odpowiadające algebrze Boole'a. W rezultacie bardzo silnie rozpowszechniło się przekonanie, iż ludzki umysł to niezwykle złożony *algorytmiczny* system obliczeniowy. Zjawiska mentalne powstają w wyniku działania połączonych ze sobą zespołów neuronów, które przesyłają potencjały elektrochemiczne wzdłuż struktur komórkowych, zwanych aksonami, poprzez połączenia zwane *synapsami*. Entuzjaści takiego stanowiska posługują się często hasłem: *Umysł ma się do mózgu tak, jak program do struktury fizycznej komputera.*[8] Innymi słowy, umysł to *software* realizowany przy pomocy niezwykle skomplikowanych obliczeń przez *hardware*, którym jest mózg. Stanowisko to nazywane jest także *funkcjonalizmem komputerowym.*[9]

Dla przedstawicieli tej koncepcji naturalną koleją rzeczy jest możliwość stworzenia odpowiedniego programu symulującego świadomy umysł. Program taki oparty musi być jedynie na odpowiednich regułach syntaktycznych, które wytworzą w nim semantykę i w ten sposób umożliwią mu *rozumienie*. Program zbudowany w oparciu o odpowiednią syntaktykę można poddać testowi Turinga[10] i próbować dowieść, że

---

[4] P. M. Churchland, *Scientific Realism and the Plasticity of Mind*, Cambrige 1979, Cambridge University Press.

[5] G. Ryle, *Czym jest umysł?,* przeł. W. Marciszewski, Warszawa 1970, PWN.

[6] J. J. C. Smart, *Doznania a procesy mózgowe*, przeł. B. Chwedeńczuk, [w:] *Filozofia umysłu*, red. B. Chwedeńczuk, Warszawa 1995, Spacja.

[7] H. Putnam, *Reason, Truth and History*, Cambridge 1981, Cambridge University Press.

[8] J. R. Searle, *Umysł na nowo odkryty,* przeł. T. Baszniak, Warszawa 1999, PIW, s. 263.

[9] Zob. np. B. Korzeniewski, *Od neuronu do (samo)świadomości*, Warszawa 2005, Prószyński i S-ka.

[10] A. Turing, *Maszyna licząca a inteligencja*, przeł. M. Szczubiałka, [w:] *Filozofia umysłu*, red. B. Chwedeńczuk, Warszawa 1995, Spacja.

jego działanie implikuje *rozumienie*. W swoim słynnym argumencie Chińskiego Pokoju,[11] amerykański filozof John Rogers Searle stara się natomiast wykazać, że operowanie zbiorem reguł formalnych w rzeczywistości nie powoduje *rozumienia.*[12] Searle jest zwolennikiem słabej (ostrożnej) wersji *AI*, która nie zakłada, że umysł to program wykonywany w mózgu, ale dopuszcza symulowanie procesów mentalnych przez procedury obliczeniowe. Wspominając o Searle'u warto zwrócić także uwagę na sugerowane przez niego rozwiązanie problematu psychofizycznego zwane *naturalizmem biologicznym.*[13] Twierdzi on bowiem, iż zjawiska mentalne są emergentnym skutkiem procesów neuronowych. Świadomość w takim ujęciu jest efektem holistycznej pracy mózgu.

Współcześnie praktycznie cała filozofia umysłu koncentruje się wokół kognitywistyki (*cognitive science*), która jest interdyscyplinarną nauką, wykorzystującą osiągnięcia takich dziedzin jak neuronauki (*neuroscience*), psychologia, lingwistyka, nauki komputerowe i inne dyscypliny wiedzy.[14] Badania te bazują na nowoczesnych technikach neuroobrazowania takich jak funkcjonalny rezonans magnetyczny (fMRI) czy emisja pozytronowa (PET), na komputerowym modelowaniu procesów poznawczych — oraz obficie czerpią z naukowych opisów klinicznych przypadków uszkodzeń mózgu (Oliver Sacks).[15] Ze względu na swój redukcjonistyczny charakter, silna hipoteza sztucznej inteligencji poddawana jest obecnie krytyce przy wykorzystaniu argumentacji, iż rozumienie opiera się na procesach *niealgorytmicznych*, zdolnych do nadawania strukturom syntaktycznym odpowiednich znaczeń. Poza tym, jednym z największych problemów dla nauk kognitywnych i filozofii umysłu jest pytanie o „lokalizację" świadomości. Czy odpowiedzialny jest za nią konkretny moduł mózgu czy też jest ona holistycznym produktem jego pracy? Jak to się dzieje, że nieświadome neurony są w stanie wytworzyć świadomy umysł?

Z odpowiedzią na pytania dotyczące genezy fenomenów umysłu i świadomości może przyjść zastosowanie osiągnięć mechaniki kwantowej do opisu procesów rozgrywających się na poziomie neuronowym. Przykładami naukowców podążających tym tropem są Max Tegmark,[16] Stuart Hameroff[17] oraz Stuart Kauffman.[18] Wczesne

[11] J. R. Searle, *Umysły, mózgi i programy,* przeł. B. Chwedeńczuk, [w:] *Filozofia umysłu*, red. B. Chwedeńczuk, Warszawa 1995, Spacja.

[12] Zob. np. J. Kloch, *Świadomość komputerów?: argument „Chińskiego Pokoju" w krytyce mocnej sztucznej inteligencji według Johna Searle'a,* Tarnów 1996, Biblos.

[13] J. R. Searle, *Umysł na nowo odkryty*, przeł. T. Baszniak, Warszawa 1999, PIW.

[14] Zob. np. W. Duch, *Czym jest kognitywistyka?* [w:] *W. Duch*, <http://www.fizyka.umk.pl/~duch/cog-book/kognitywistyka.htm>.

[15] Zob. np. O. Sacks, *Mężczyzna który pomylił swoją żonę z kapeluszem*, przeł. B. Lindenberg, Poznań 1996, Zysk i S-ka.

[16] M. Tegmark*, The importance of quantum decoherence in brain processes*, [w:] *Cornell University Library, Arxiv.org*, < http://arxiv.org/abs/quant-ph/9907009>.

[17] S. Hameroff, S. Hagan, J. A. Tuszyński, *Quantum computation in brain microtubules? Decoherence and biological feasibility*, [w:] *Cornell University Library, Arxiv.org*, <http://arxiv.org/abs/quant-ph/0005025>.

[18] S. A. Kauffman, *Reinventing the sacred*, New York 2008, Basic Books.

stadium tych badań powoduje jednak, iż kwantowe koncepcje umysłu posiadają wysoce spekulatywny charakter, a nawet, jak to ma miejsce w przypadku Kauffmana, bazują na przestarzałych interpretacjach mechaniki kwantowej oraz ewidentnych błędach w rozumieniu podstawowych pojęć filozoficznych takich jak na przykład *przyczynowość*. Wykorzystanie metod kwantowych zaowocowało również niedawno próbą stworzenia fizycznej teorii subiektywnych stanów mentalnych, mającej na celu obiektywizację danych kognitywnych, otrzymywanych na bazie świadectw badanych podmiotów.[19]

W tym roku mija dwadzieścia lat od ukazania się *Nowego umysłu cesarza*[20] — książki, w której jej autor, znany brytyjski matematyk i fizyk, Roger Penrose, po raz pierwszy zaprezentował niezwykle oryginalną koncepcję funkcjonowania ludzkiego umysłu. Oryginalność jego podejścia polega na tym, iż koncepcja ta wyłania się w szerszym kontekście poszukiwań nowej teorii fizycznej, unifikującej teorię względności z mechaniką kwantową — *kwantowej grawitacji*. Penrose wysuwa bowiem hipotezę wzajemnego współistnienia trzech światów — świata *matematyki*, świata *fizyki* oraz świata *umysłu*. Nowa uogólniona teoria będzie te trzy światy scalała w ramach jednego, zunifikowanego formalizmu matematycznego. Zdając sobie sprawę z nowatorstwa sugerowanego przez siebie rozwiązania, wzajemne relacje między matematyką, materią (światem fizycznym) a umysłem, Penrose określa jako *tajemnice*. Choć wydawać by się mogło, iż jego koncepcja umysłu otrzymała już należną porcję zmasowanej krytyki, zmuszając go ostatecznie do ponownego sprecyzowania swoich tez w *Cieniach umysłu*,[21] ponowne przedyskutowanie tej koncepcji pozwoli uwzględnić szereg nowych osiągnięć w teorii umysłu, a także włączyć do analizy refleksje Penrose'a, zawarte w jego ostatnio opublikowanym monumentalnym dziele *Droga do rzeczywistości*.[22] Ważniejszym elementem niniejszej pracy będzie jednak zwrócenie uwagi na to, jak prezentowane przez Penrose'a *filozoficzne* stanowisko matematycznego platonizmu wpływa na kształtowanie się formułowanych przez niego teorii fizycznych, a w szczególności próba wskazania, co prowadzi do ostatecznego postulowania ich *niealgorytmicznego* charakteru. Postaramy się pokazać istotną rolę „filozofii w kontekście nauk" w docieraniu do najgłębszych inspiracji i motywów, które towarzyszą rozwojowi współczesnej nauki. Mogło by się wręcz wydawać, iż kontrowersyjność sugerowanych przez Penrose'a rozwiązań wielu problemów współczesnej fizyki to wina uprawianej przez niego filozofii, ale, jak stwierdza ks. Michał Heller, *warto się dobrze zastanowić, zanim się z Penrosem nie zgodzimy.*[23]

---

[19] S. Lee, *Towards a Physical Theory of Subjective Mental States*, [w:] *Cornell University Library, Arxiv.org,* <http://arxiv.org/pdf/0712.2873v1>.

[20] R. Penrose, *Nowy umysł cesarza*, przeł. P. Amsterdamski, Warszawa 2000, PWN.

[21] R. Penrose, *Cienie umysłu*, przeł. P. Amsterdamski, Poznań 2000, Zysk i S-ka.

[22] R. Penrose, *Droga do rzeczywistości*, przeł. J. Przysława, Warszawa 2006, Prószyński i S-ka.

[23] M. Heller, *Kosmologia kwantowa*, Warszawa 2001, Prószyński i S-ka.

KONCEPCJA TRZECH ŚWIATÓW PENROSE'A

Dokładniejsze określenie miejsca, jakie zajmuje umysł w ontologii Penrose'a, oraz perspektywy, jaką w jej świetle uzyskuje problem *mind-body*, wymaga przybliżenia specyfiki tej ontologii ze szczególnym uwzględnieniem relacji *świat fizyczny-umysł* oraz *umysł-matematyka*. Pierwszy z trzech światów w ontologii Penrose'a to świat platońskich idei matematycznych. Byty matematyczne takie, jak liczby naturalne, maszyny Turinga, równania Einsteina czy też zbiory Mandelbrota, istnieją w nim realnie i niezależnie od poznającego podmiotu. Innymi słowy, świat ten nie jest stwarzany przez matematyków, lecz ma charakter obiektywny i może być odkrywany przez ludzki umysł, co nadaje sens pojęciu obiektywnej prawdy matematycznej. Taki pogląd w filozofii matematyki nazywany jest *platonizmem matematycznym.*[24] Jako zdeklarowany platonik, Penrose twierdzi, że:

> Obiektywne pojęcia matematyczne należy uważać za byty ponadczasowe, a nie za powołane do życia z chwilą zauważenia ich po raz pierwszy przez człowieka.[25]

Do świata tego należą także idee takie, jak dobro, prawda i piękno.

Z pozaprzestrzennego i pozaczasowego świata matematyki wyłania się świat fizyczny, który opisywany jest przez zmatematyzowane nauki szczegółowe. Relacja ta sanowi dla Penrose'a (a także dla innych badaczy[26]) podstawę do wysunięcia hipotezy o *matematyczności* Wszechświata, zgodnie z którą jego strukturą z wielką dokładnością rządzą „bezczasowe prawa matematyczne".[27] Istnieje także trzeci świat — świat umysłu, który obejmuje swoim zasięgiem fenomeny takie jak świadomość, rozumienie czy inteligencja. W opinii Penrose'a świat umysłu jest emergentny wobec świata fizycznego, co oznacza, iż wszystkie zjawiska mentalne znajdują swoje ostateczne uzasadnienie w świecie fizyki (fizykalizm).

Z trzema światami związane są trzy tajemnice: w jaki sposób świat fizyczny wyłania się z idealnego świata matematyki?, jak w świecie fizycznym powstają świadome umysły? i wreszcie: co zapewnia naszym umysłom dostęp do pozaczasowego i pozaprzestrzennego świata idei? Ważnym elementem koncepcji Penrose'a, który dla podkreślenia jego hipotetyczności określa on mianem przesądu (ang. *prejudice*), jest przyjęcie, iż każdy następny świat *w całości* wyłania się jedynie z części poprzedniego. W efekcie świat fizyczny wyłania się tylko z części bogatego świata matematyki, a świat umysłu wyłania się natomiast jedynie z niektórych struktur świata fizycznego, odpowiedzialnych za budowę mózgu. Jeśli chodzi natomiast o świat platoński, to tylko niewielka część aktywności umysłowej dotyczy absolut-

---

[24] Zob. np. K. Wójtowicz, *Platonizm matematyczny*, Tarnów 2001, Biblos.

[25] R. Penrose, *Droga…*, s. 17.

[26] Zob. np. M. Heller, *Filozofia i wszechświat*, Kraków 2006, Universitas.

[27] R. Penrose z udz. A. Shimony, N. Cartwrtight, S. Hawking, *Makroświat, mikroświat i ludzki umysł*, przeł. P. Amsterdamski, red. M. Longair, Warszawa 1997, Prószyński i S-ka, s. 18.

nych idei. Penrose twierdzi ponadto, że nie istnieją prawdy matematyczne należące do świata platońskiego, które nie byłby dostępne dla świadomego umysłu.

W ten sposób możliwa staje się aktywność poznawcza, która wedle Penrose'a polega na odnajdywaniu w świecie platońskim odpowiednich konstrukcji matematycznych, opisujących rzeczywistość świata fizycznego:

> Im lepiej rozumiemy świat fizyczny, im głębiej poznajemy prawa natury, tym bardziej wydaje się nam, że świat fizyczny gdzieś wyparowuje i pozostaje nam tylko matematyka. Im głębiej rozumiemy prawa fizyki, tym dalej wkraczamy w świat matematyki i matematycznych pojęć.[28]

Wedle Penrose'a to świat matematyki jest najbardziej pierwotną podstawą dla rzeczywistości, a świat fizyczny nie jest, jak twierdził Platon, jedynie niedoskonałą kopią świata idei. Penrose wyraźnie mówi, że

> (…) z tej trójki świat matematycznych idei Platona jest, w pewnym sensie, najbardziej prymitywny, ponieważ matematyka stanowi rodzaj konieczności, magicznie wyczarowując siebie siłą samej logiki.[29]

Każdy kolejny świat wyłaniający się z części poprzedniego jest coraz bogatszy. Tak więc najbardziej złożony i najbogatszy jest świat umysłu. Penrose nie uważa, że poszczególne światy są od siebie oddzielone, tak jak chciał tego Platon (koncepcja *horismos*), lecz stanowią one być może część tajemniczej i znacznie bogatszej całości. Co więcej, Penrose uznaje, że zrozumienie jednego z trzech obszarów nie jest możliwe bez odwołania się do dwóch pozostałych. Powinny one zostać zunifikowane w ramach formalizmu nowej teorii fizycznej, której jedną z podstawowych cech będzie niealgorytmiczność. Niealgorytmiczność (nieobliczalność) stanowi z kolei istotne ogniwo, spajające ze sobą świat umysłu ze światem matematyki. Penrose uważa, że z uwagi na niealgorytmiczność wielu rozumowań matematycznych *rozumienie* wiąże się z wglądem w świat bytów matematycznych, natomiast sama świadomość nie powstaje na skutek obliczeń wykonywanych przez mózg, lecz jest skutkiem nieobliczalnych procesów zachodzących na poziomie neuronowym.

ZARYS ARGUMENTU PENROSE'A

Argumentacja Penrose'a łączy w sobie kwestie natury logiczno-matematycznej, fizycznej, biologicznej oraz psychologicznej. Warto zaznajomić się z ogólnym schematem, na którym Roger Penrose buduje swoją koncepcję umysłu:

1. Na początek prosta zjawiskowa obserwacja — ludzkie rozumienie posiada charakter *niealgorytmiczny*, o czym można przekonać się, obserwując np. grę w szachy.

2. Krok drugi — próba formalizacji powyższego zagadnienia za pomocą maszyny Turinga i zastosowaniu do niej twierdzenia Gödla o niezupełności systemów for-

[28] Tamże, s. 18-19.
[29] R. Penrose, *Droga*…, s. 990.

malnych, opartych na arytmetyce liczb naturalnych (nierozstrzygalność problemu stopu). Wniosek: wgląd matematyczny nie może być zakodowany w procedurze algorytmicznej, jeśli istnieje pewność co do jej poprawności.

3. Niealgorytmiczność wglądu matematycznego implikuje niealgorytmiczność ludzkiej świadomości jako takiej (główny punkt argumentacji Penrose'a).

4. Istnieje przypuszczenie, że unifikacja mechaniki kwantowej z teorią względności wykaże niealgorytmiczny charakter kwantowej grawitacji. W szczególności procesem takim będzie grawitacyjnie indukowana procedura redukcji wektora falowego **R**.

5. Wychodząc od molekularnej struktury mózgu (tj. struktury neuronowej), można pokazać, iż mikrotubule, stanowiące jeden z fundamentalnych składników strukturalnych neuronów, są na tyle izolowane od otoczenia, iż w ich obrębie mogą wystarczająco długo utrzymywać się kwantowe stany splątane. Jeśli obejmą one znaczne obszary mózgu, wówczas świadomość przyjmie charakter nielokalny.

6. Konkretne akty świadomości mają miejsce wtedy, gdy następuje obiektywna redukcja stanu splątanego (**OR**), w którym pozostają biologicznie aktywne cząsteczki tubuliny, prowadząca do wygenerowania odpowiedniej obserwabli w reżimie klasycznym. W konsekwencji niealgorytmiczność procedury **OR** powoduje niealgorytmiczność świadomości.

Warto zwrócić uwagę, iż rozumowanie Penrose'a rozpada się na dwa zasadnicze segmenty: (1)-(3): czysto makroskopowa (zjawiskowa) obserwacja przejawów niealgorytmiczności myślenia, uściślona formalnym argumentem o nierozstrzygalności w kontekście arytmetyk opartych na liczbach naturalnych. (4)-(6): konstrukcja molekularnego (mikroskopowego) modelu mózgu, pozwalająca wykazać, że emergentne struktury biologiczne mogą być w skali makro odpowiedzialne za niealgorytmiczne zjawiska fizyczne, generujące świadomość. Pouczającą obserwację stanowi również fakt, iż relacja tych dwóch segmentów jest analogiczna do relacji, w jakiej pozostają między sobą termodynamika fenomenologiczna i statystyczna. Podczas gdy termodynamika fenomenologiczna bada procesy cieplne jedynie od strony ich makroskopowych przejawów, to wychodząc od założeń co do drobinowej struktury ciał na poziomie mikro, termodynamika statystyczna wyprowadza z nich odpowiednie własności makroskopowe rejestrowane fenomenologicznie.

DLACZEGO MYŚLENIE NIE JEST ALGORYTMICZNE?

Jedną z oczywistych trudności w analizie algorytmiczności ludzkiego myślenia jest rozbieżność pomiędzy potocznym i opisowym charakterem takich pojęć jak *wgląd*, *rozumienie* czy też *świadomość* a ścisłym, matematycznym pojęciem *algorytmiczności*. Roger Penrose nie widzi jednak konieczności precyzyjnego określania tych pojęć, ponieważ z punktu widzenia matematyka (a taki Penrose prezentuje) ważniejsze i *de facto* wystarczające do analizy zagadnienia są zachodzące między

nimi relacje.[30] I tak *rozumienie,* zwane również *wglądem*, nie jest możliwe bez *świadomości*, podobnie jak nie jest możliwa *inteligencja* bez *rozumienia*. Jeśli więc będzie można wykazać, iż rozumienie w matematyce jest niealgorytmiczne, to przy założeniu, że rozumienie wykracza poza zakres matematyki, wniosek ten będzie można rozszerzyć i postawić tezę, że ludzka świadomość jako całość posiada niealgorytmiczną naturę.

Penrose rozpoczyna swoje dość złożone uzasadnienie powyższej tezy od wyróżnienia czterech stanowisk na temat możliwych związków między zjawiskami mentalnymi a *obliczeniami.*[31] Zgodnie ze stanowiskiem **A**, które nazywane jest *silną sztuczną inteligencją* (*strong AI*) lub *funkcjonalizmem komputerowym,* wszelkie zjawiska mentalne (w tym również świadomość) polegają na wykonywaniu obliczeń na bazie odpowiednio złożonego algorytmu. Stanowisko **B** zakłada natomiast, iż możliwe jest komputerowe symulowanie zjawisk rozgrywających się na poziomie neuronowym, ale nie wywołuje ono nigdy świadomości. Zjawiska mentalne związane są nieodłącznie z fizyczną strukturą mózgu i zachodzącymi w nim procesami biochemicznymi, natomiast sama komputerowa symulacja połączeń neuronowych nie generuje świadomości. Zwolennikiem takiego stanowiska jest wspominany już wcześniej John Searle. Stanowisko **C**, za którym opowiada się Penrose, zakłada, że za powstawanie świadomości odpowiedzialne są procesy fizyczne zachodzące w mózgu, których nie da się jednak symulować obliczeniowo. Zgodnie ze stanowiskiem **D** nie należy poszukiwać wyjaśnienia zjawisk mentalnych z pomocą jakichkolwiek metod naukowych. Warto również zaznaczyć, iż w stanowisku **C**, którego reprezentantem jest Penrose, wyróżnić można dwie wersje: *słabą* i *silną*. Wedle pierwszej nieobliczalne procesy można opisać za pomocą znanych dotychczas praw fizycznych. Natomiast według silnej wersji **C** w dotychczas znanych teoriach fizycznych nie występują żadne procesy nieobliczalne — jak pisze Penrose:

> nasza znajomość fizyki nie wystarcza do wyjaśnienia świadomości, a zatem musimy wyjść poza znane prawa fizyczne.[32]

Pojawia się jednak pytanie: co Penrose rozumie przez obliczalność i nieobliczalność?

By wyjaśnić pojęcie obliczalności, należy odwołać się do pojęcia *algorytmu*, które Penrose traktuje jako synonim terminu *obliczenia.*[33] Jako jeden z przykładów podaje on algorytm Euklidesa, który jest systematyczną procedurą, służącą do odnaj-

---

[30] R. Penrose z udz. A. Shimony, N. Cartwrtight, S. Hawking, *Makroświat…,* s. 105.

[31] Tamże, s. 105-106.

[32] Tamże, s. 106.

[33] Penrose zwraca uwagę na fakt, iż niektórzy autorzy nie stosują pojęcia algorytmu do obliczeń o oddolnej organizacji, takich jak np. sieci neuronopodobne. Sam Penrose pojęcie algorytmu stosuje szeroko, a więc zarówno do odgórnych, jak i oddolnych procedur obliczeniowych, ponieważ obydwie mogą być realizowane za pomocą zwykłego komputera. Więcej na ten temat znaleźć można w: R. Penrose, *Cienie…,* s. 36-38.

dywania największego dzielnika dla dowolnej pary liczb. W języku potocznym przez algorytm najczęściej rozumie się:

> (…) efektywną procedurę, której zastosowanie pozwala osiągnąć zamierzony cel (znaleźć rozwiązanie konkretnego problemu) w skończonej liczbie kroków.[34]

Mimo wielu procedur tego typu znanych w matematyce na przestrzeni dziejów, samo pojęcie *algorytmu* zostało precyzyjnie zdefiniowane dopiero w XX wieku. Najbardziej znaną ścisłą definicją algorytmu jest funkcjonalny opis oparty na *maszynie Turinga*. Pojęcie to zostało stworzone przez angielskiego matematyka Alana Turinga w odpowiedzi na pytanie Dawida Hilberta dotyczące możliwości istnienia algorytmicznej procedury rozstrzygania problemów matematycznych (tzw. *Entscheidungsproblem*). Maszyna Turinga nie jest urządzeniem fizycznym, lecz wyidealizowanym obiektem matematycznym, posiadającym skończony, dyskretny zbiór możliwych stanów wewnętrznych. Mimo skończonego zbioru tych stanów, maszyna ta może korzystać z nieograniczonej pamięci zewnętrznej i podawać wyniki dowolnego rzędu. Maszyna Turinga, która należy według Penrose'a do platońskiego świata bytów matematycznych, jest idealizacją współczesnego komputera. Zdaniem Penrose'a,

> (…) Pojęcie maszyny Turinga (lub równoważne) rzeczywiście określa, co z matematycznego punktu widzenia rozumiemy przez procedurę algorytmiczną (konstruktywną, rekurencyjną lub mechaniczną).[35]

Wynika z tego, iż w przeciwieństwie do procesów obliczalnych, procesy nieobliczalne nie mogą być realizowane przez żaden rzeczywisty komputer czy też jego wyidealizowaną wersję, a więc przez maszynę Turinga.

Wedle postulowanej przez Penrose'a *mocnej* wersji stanowiska **C**, wszystkie znane nam prawa fizyczne są w pełni obliczalne, a więc mogą być realizowane przez maszyny Turinga. Do obliczalnych należą zatem prawa rządzące układem kwantowym (równanie Schrödingera), mechaniką klasyczną i teorią względności, a także zjawiska chaosu deterministycznego. Aby zatem wskazać przykłady *rozumienia* o charakterze niealgorytmicznym, Penrose odwołuje się do świata matematyki. Przykładem zagadnienia matematycznego, którego nie da się rozwiązać za pomocą żadnego algorytmu jest obszernie analizowany przez niego problem pokrycia płaszczyzny Euklidesa skończoną liczbą różnokształtnych płytek (problem parkietażu). Penrose podaje także „ćwiczebny model wszechświata", w którym czas jest wielkością dyskretną, a jego stan w określonej chwili opisywany jest przez dwa zbiory polikwadratów. Wszechświat ten ewoluuje w pełni deterministycznie, lecz zarazem nieobliczalnie, gdyż, zgodnie z twierdzeniem Bergera,[36] nie istnieje algorytm, dzięki któremu można by symulować tę procedurę na komputerze. Przykład tego ćwiczebnego

---

[34] J. Życiński, *Nieobliczalny wszechświat Penrose'a*, [w:] *Granice racjonalności,* Warszawa 1993, PWN, s. 221-222.

[35] R. Penrose, *Nowy umysł…*, s. 66.

[36] R. Penrose z udz. A. Shimony, N. Cartwrtight, S. Hawking, *Makroświat…*, s. 122.

modelu wskazuje, iż, jak pisze Penrose: „(…) determinizm to coś innego niż obliczalność".[37]

Kluczowym argumentem Penrose'a na rzecz niealgorytmiczności wglądu matematycznego jest twierdzenie Gödla. Twierdzenie to powstało jako odpowiedź na program Hilberta, którego celem było zbudowanie pewnych i niewzruszalnych podstaw dla matematyki.[38] Dotyka ono dwóch ważnych dla matematyki kwestii, mianowicie prawdy i dowodu. Kurt Gödel wykazał bowiem, że w każdym opartym na arytmetyce systemie formalnym znajdują się poprawnie sformułowane zdania, których wartość logiczną można orzec, ale nie da się jej w ramach tego systemu udowodnić. Problemu tego nie rozwiązuje także „dokładanie" nowych aksjomatów, gdyż i tak pojawiać się będą zdania prawdziwe, lecz niedowodliwe w danym systemie aksjomatów. Ponadto, zgodnie z tzw. drugim twierdzeniem Gödla, dowód niesprzeczności danego systemu formalnego wymaga sięgnięcia po środki logiczne nienależące do tego systemu. Jak pisze M. Heller:

> (…) jeżeli da się udowodnić niesprzeczność systemu sformalizowanego (…), to można to zrobić jedynie przez odwołanie się do bogatszego systemu sformalizowanego.[39]

Skoro pewne poprawnie zbudowane zdania nie dają się dowieść na gruncie systemu formalnego, to skąd wiadomo, że są prawdziwe? Penrose uważa, że pojęcie prawdy matematycznej wykracza poza formalizm, a jej odkrywanie możliwe jest dzięki *zasadzie refleksji*, zgodnie z którą:

> (…) dokonując refleksji nad znaczeniem systemu aksjomatów i reguł wnioskowania i przekonując się, że rzeczywiście dostarczają one poprawnych metod dochodzenia do prawdziwych wniosków, można dojść do dalszych prawdziwych stwierdzeń matematycznych, których nie sposób wydedukować z aksjomatów za pomocą przyjętych reguł wnioskowania.[40]

Do odkrycia prawdy potrzebny jest zatem według Penrose'a pewien rodzaj bezpośredniego *wglądu* w świat bytów matematycznych. Operacje formalne nie są wystarczającym środkiem do ustalenia prawdziwości zdań. Penrose uważa, że dzięki *zasadom refleksji* wykraczającym poza formalizm matematycy mają możliwość rozważania takich bytów jak np. zbiory nieskończone. Stanowisko takie wymaga przyjęcia wspominanego już wcześniej platonizmu matematycznego, a więc poglądu, wedle którego obiekty matematyczne istnieją realnie i pozaczasowo, a pojęcie prawdy matematycznej ma absolutny charakter i wymaga od matematyka wglądu w ów istniejący absolutnie świat idei matematycznych.

Rozważanie kwestii *rozumienia* w matematyce można zdaniem Penrose'a ekstrapolować na *rozumienie* w każdej innej dziedzinie działalności umysłu. Wedle niego żaden typ aktywności umysłowej, który wymaga rozumienia, nie daje się ująć w al-

[37] R. Penrose, *Nowy umysł...*, s. 123.

[38] Zob. np. R. Murawski, *Filozofia matematyki*, Warszawa 2001, PWN.

[39] M. Heller, *Przeciw fundacjonizmowi*, [w:] *Filozofia i wszechświat*, Kraków 2006, Universitas, s. 87.

[40] R. Penrose, *Nowy umysł...*, s. 132.

gorytm. Przykład matematyki pokazuje zdaniem Penrose'a, że myślenie ma charakter zdecydowanie niealgorytmiczny i wymaga pewnego rodzaju *wglądu*. Nieobliczeniowe ujęcie rozumienia ekstrapolowane jest przez Penrose'a także na inne zjawiska mentalne, w tym samą świadomość.

KWANTOWE PODSTAWY MYŚLENIA

Przedstawione powyżej czysto zjawiskowe argumenty na rzecz niealgorytmiczności ludzkiego myślenia dostarczają szczegółowego wyjaśnienia pierwszego segmentu (punkty 1-3) w uprzednio naszkicowanym zarysie Penrose'a koncepcji umysłu. W ten sposób wyczerpująco opisana zostaje również relacja zachodząca pomiędzy *światem umysłu* i *światem matematyki*. Pełna analiza funkcjonowania umysłu wymaga obecnie sięgnięcia do poziomu mikro, co Penrose realizuje poprzez budowę modelu umysłu, gdzie makroskopowo stwierdzoną niealgorytmiczność myślenia warunkować będą procesy kwantowe. Model ten odgrywa w koncepcji Penrose'a kluczową rolę dla dokładniejszego scharakteryzowania relacji, jaka zachodzi pomiędzy światem fizycznym a światem umysłu. Warto również w tym miejscu jeszcze raz zaakcentować, iż w relacji tej wyraża się *fizykalistyczne* stanowisko Penrose'a, równoważne poglądowi, iż wszystkie zjawiska mentalne ostatecznie redukowalne są do zjawisk fizycznych. Innymi słowy, umysł stanowi strukturę emergentną w stosunku do świata fizyki.

Podstawowym procesem fizycznym, który stanowi dla Penrose'a punkt wyjścia w konstrukcji mikroskopowego modelu mózgu jest kwantowa procedura *redukcji wektora falowego* **R**. Z uwagi na fakt, iż zagadnienie to omawiane było w kilku odrębnych publikacjach, na obecne potrzeby wystarczy ograniczyć się do kilku niezbędnych uwag.[41] Ewolucja czasowa stanów kwantowych, opisanych wektorem falowym, która odbywa się zgodnie z deterministycznym równaniem Schrödingera, zwana jest procedurą unitarnej ewolucji **U** tego wektora. Podobnie jak w równaniach mechaniki klasycznej, procedura ta jest całkowicie deterministyczna, liniowa oraz symetryczna (odwracalna) ze względu na czas. Sytuacja taka ma miejsce dopóty, dopóki na układzie kwantowym nie zostanie dokonany pomiar, który wiąże się ze wspomnianą już powyżej procedurą redukcji wektora falowego **R**. W rezultacie urządzenie pomiarowe wskaże wynik — obserwablę, będącą wartością własną odpowiedniego operatora kwantowo-mechanicznego, wybraną z właściwym sobie prawdopodobieństwem spośród wszystkich, które wchodzą w skład stanu kwantowego, opisanego równaniem. W odróżnieniu od procedury **U**, procedura **R** jest *indeterministyczna*, *nieliniowa* oraz *nieodwracalna* w czasie. Oczywista niekompatybilność tych dwóch procedur stanowi główne źródło problemów mechaniki kwantowej, co potocznie określa się mianem *problemu pomiaru* lub, jak woli Penrose, *paradoksu pomiaru*. Wszystkie podejmowane wysiłki interpretacyjne mechaniki kwan-

[41] Zob. np. W. Grygiel, *Zobaczyć kota Schrödingera*, Tarnów 2008, Biblos, s. 131-152.

towej stawiają sobie za cel rozwiązanie tego paradoksu albo poprzez jego wchłonięcie wewnątrz ogólniejszego formalizmu (np. *spójne historie kwantowe*), albo potraktowanie jako obiektywny proces fizyczny — za czym ewidentnie opowiada się Penrose. Tak w jednym, jak i w drugim przypadku, problem pomiaru dotyka bardzo istotnego zagadnienia *emergencji* świata makroskopowego z poziomu kwantowego, ponieważ redukcja wektora falowego zachodzi na skutek oddziaływania badanego układu kwantowego z makroskopowym otoczeniem.

Przeglądając całą plejadę różnych interpretacji mechaniki kwantowej, odnosi się słuszne wrażenie, iż problemy tej teorii tkwią dużo głębiej niż tylko w prostym braku kompatybilności pomiędzy kwantowym formalizmem przestrzeni Hilberta a językiem fizyki klasycznej, w kontekście którego analizowane są dane pomiarowe. Stąd też jasna wydaje się opinia Penrose'a, iż współczesna mechanika kwantowa uwikłana jest w fundamentalne problemy, które uzyskają dopiero swoje pełne rozwiązanie w obrębie przyszłej teorii kwantowej grawitacji, łączącej kwantową strukturę mikroświata ze strukturą Wszechświata jako całości, wynikającą z ogólnej teorii względności.

W przeciwieństwie do zdania większości wybitnych współczesnych fizyków takich, jak chociażby Stephen Hawking, Roger Penrose twierdzi, iż nie jest możliwe sformułowanie takiej interpretacji mechaniki kwantowej, która doprowadziłaby do zniwelowania wszystkich problemów tej teorii i dlatego niezbędne jest takie uogólnienie jej sformułowania, które umożliwi skwantowanie grawitacji. Nieliniowość procedury redukcji wektora falowego **R** zwiastuje zdaniem Penrose'a, iż rzeczywistość kwantowa w swej prawdziwej — a jeszcze nie do końca zbadanej odsłonie — jest *nieliniowa*. Z uwagi na fakt, iż koreluje to z nieliniowym charakterem ogólnej teorii względności, istnieje nadzieja, że w swoim nowym kształcie mechanika kwantowa pozwoli w naturalny sposób skwantować pole grawitacyjne. W takich warunkach obie dotychczas niekompatybilne procedury **U** i **R** zostaną zunifikowane w obrębie jednego formalizmu, a redukcja wektora falowego okaże się skutkiem działania *kwantowej grawitacji*.

Potraktowanie procedury **R** jako rzeczywistego procesu fizycznego wynika dodatkowo z utrzymywanego przez Penrose'a stanowiska, iż wektor falowy Ψ jest obrazem realnie istniejącego stanu fizycznego badanego układu.[42] W celu zilustrowania sugerowanej przez siebie hipotezy co do mechanizmu grawitacyjnej redukcji wektora falowego Penrose odwołuje się do różnic, jakie mogą wystąpić w geometriach czasoprzestrzeni, odpowiadającym różnym stanom, wchodzącym w skład ich liniowej kombinacji. Różnice te są na tyle głębokie, iż skutecznie uniemożliwiają współistnienie tych czasoprzestrzeni w obrębie jednego stanu splątanego i dlatego prowadzą do redukcji wektora falowego.[43] Co więcej, szybkość redukcji zależy od masy obiektów, których stany brane są pod uwagę tak, że odwrotność wartości energii,

[42] W. Grygiel, *Interpretacje mechaniki kwantowej jako ontologie mikroświata*, „Logos i Ethos", 1(24) 2008, s. 59-72.
[43] R. Penrose, *Nowy umysł…,* s. 132.

potrzebnej na przezwyciężenie siły grawitacji przy rozsuwaniu tych obiektów na pewną odległość, mierzy *połowiczny czas życia* stanu splątanego.[44]

Próbując obecnie wykorzystać powyższy schemat do obiektów o znaczeniu biologicznym, Penrose proponuje zamiast pojedynczych punktów masowych w stanie splątanym rozważyć różniące się przestrzennym rozkładem masy możliwe konformacje molekuł odpowiadających tym obiektom. W przypadku, gdy molekuły takie byłyby dostatecznie izolowane od środowiska, istnieje prawdopodobieństwo, iż kwantowe superpozycje ich różnych konformacji mogłyby utrzymywać się nawet przez dłuższy czas.[45] W ten sposób procedura redukcji wektora falowego **R** mogłaby dotyczyć stanów molekuł o rozmiarach rzędu mikronów lub nawet milimetrów, dzięki czemu stany splątane mogłyby utrzymywać się nie tylko w skali pojedynczych cząstek elementarnych, ale znacznie większych obiektów, takich przykładowo, z jakich zbudowany jest mózg. Redukcję wektora falowego w takich warunkach Penrose nazywa mianem redukcji obiektywnej i oznacza skrótem **OR**. Przytaczając dodatkowo szereg przykładów obserwowanych zjawisk nadprzewodnictwa w układach biologicznych, Penrose stara się uzasadnić możliwość współistnienia wielu dużych molekuł w jednym, spójnym stanie kwantowej superpozycji. W rezultacie wysnuwa on przypuszczenie, iż procedura **OR**, która może angażować nawet znaczniejsze obszary mózgu, będzie posiadała *niealgorytmiczny* charakter, co odpowiada przedstawionej we wcześniejszej części niniejszej pracy czysto obserwacyjnej konkluzji o niealgorytmiczności ludzkiego myślenia. Pełne uzasadnienie teorii opisującej **OR** wymaga jednak stworzenia nowej teorii *kwantowej grawitacji*, dlatego też nie dziwi podtytuł *Cieni umysłu — Poszukiwanie brakującej teorii świadomości*. Teoria ta będzie, zdaniem Penrose'a, miała niealgorytmiczny charakter. Na poparcie takiej hipotezy, zwraca on w *Cieniach umysłu* uwagę na osiągnięcia Roberta Gerocha i Jamesa Hartle'a (topologiczne aspekty czterowymiarowej czasoprzestrzeni)[46] oraz Davida Deutscha (zapętlone krzywe czasoprzestrzenne).[47]

W swoim ostatnim dziele *Droga do rzeczywistości* Roger Penrose szczegółowo omawia proponowaną przez siebie strategię stworzenia teorii kwantowej grawitacji na tle innych, konkurencyjnych rozwiązań.[48] W ten sposób struktura świata fizycznego w omawianym powyżej obrazie trzech światów ma szanse uzyskać ściślejsze matematyczne uzasadnienie, które bazuje na założeniu występowania procesów niealgorytmicznych w przyrodzie, zwłaszcza jeśli chodzi o strukturę i funkcjonowanie umysłu. Ponadto cechą charakterystyczną rozwoju teorii fizycznych, jasno akcentowaną przez Penrose'a wręcz jako jeden z podrozdziałów *Drogi do rzeczywistości*, jest stymulacja tego rozwoju przez podstawowe idee matematyczne (ang. *mathematically driven fundamental physics*). W szczególności, Penrose wskazuje na opraco-

[44] R. Penrose, *Cienie umysłu*…, s. 341.
[45] Tamże, s. 458.
[46] Tamże, s. 464.
[47] Tamże, s. 468.
[48] R. Penrose, *Droga…,* s. 925 i nast.

wywaną przez siebie on ponad czterdziestu lat teorię *twistorów*, u podstaw której leży bardzo elegancka matematycznie idea *liczb zespolonych*. Z jednej strony, pozwala ona odejść od klasycznego, to jest ciągłego modelu czasoprzestrzeni, opartego na liczbach rzeczywistych, a z drugiej nie wiązać się z popularnymi dziś modelami dyskretnymi. W ten sposób, jak stwierdza Penrose:

> Będzie można zauważyć, w jak nieoczekiwany i znaczący sposób teoria twistorów wchodzi w relację z ogólną teorią względności, dając jednocześnie bardzo interesujące perspektywy w obszarze kwantowej teorii pola (QFT), fizyki cząstek oraz nieliniowego uogólnienia mechaniki kwantowej.[49]

Warto również zwrócić uwagę, iż w teorii twistorów pojawia się intrygujące z punktu widzenia topologii pojęcie *ko-homologii.*[50] Pozwala ono w sposób globalny opisywać nielokalność, przez co może stanowić bardzo naturalne, choć matematycznie złożone, środowisko do opisu nielokalnych aspektów świadomości. Choć Penrose nie wypowiada tego jednoznacznie, tak ściśle zdefiniowane pojęcia matematyczne jak ko-homologie przynależą z pewnością do platońskiego świata matematyki. Z uwagi na ich wysoce nieliniowy charakter (całość niesprowadzalna do części) można również przypuszczać, iż nie da się ich przedstawić w postaci funkcji rekurencyjnych, w wyniku czego ich analiza wymagać będzie narzędzi niealgorytmicznych. Naturalnie więc znajdą się one u podstaw niealgorytmicznej fizyki, w spójny sposób traktującej zjawiska nielokalności i być może dającej głębszy wgląd w funkcjonowanie ludzkiej świadomości.

KOHERENTNE PROCESY KWANTOWE JAKO PODŁOŻE DZIAŁANIA MÓZGU

W modelu działania mózgu, jakim posługują się współcześni specjaliści z zakresu neurobiologii i kognitywistyki, uwzględnia się zjawiska występujące na poziomie neuronowej struktury tkanki mózgowej. Roger Penrose zauważa, że zwolennicy takiego stanowiska przeważnie nie negują istnienia zjawisk kwantowych, odnoszących się do działania mózgu w skali mikroskopowej. Twierdzą oni jednak, że wyjaśnienia na poziomie klasycznym są zadowalające i kompletne. Zgodnie z powszechnie znanym schematem, na poziomie neuronowym impulsy nerwowe przechodzą z centralnego ciała komórki nerwowej wzdłuż aksonu, który rozgałęzia się na włókna zakończone synapsami, gdzie przez szczeliny synaptyczne impuls przenoszony jest do dendrytów kolejnej komórki nerwowej lub bezpośrednio do jej wnętrza. Synapsy dzielą się na pobudzające i hamujące, gdyż, w zależności od konkretnego neurotransmitera, następny neuron jest pobudzany do wysłania impulsu lub przesłanie impulsu zostaje zahamowane.

---

[49] Tamże, s. 964.
[50] Tamże, s. 987.

Działanie mózgu można na bazie powyższego schematu modelować obliczeniowo.[51] Układy synaps i wagi połączeń synaptycznych są w takim modelowaniu ustalone częściowo odgórnie, a częściowo korzysta się z czynników losowych. Z uwagi na fakt, iż w rzeczywistych mózgach wagi synaps są jednak dynamiczne i podlegają zmianom, musi, zdaniem Penrose'a, istnieć odpowiedni proces, który tymi zmianami kieruje. Penrose zwraca uwagę, iż:

> w modelach sieciowych (takich, jakie przyjmuje się przy projektowaniu sztucznych sieci neuronowych) obowiązują obliczeniowe reguły modyfikowania synaps.[52]

Reguły w sieciach neuronowych zostają wybierane tak, by sprawność układu stale rosła. Kryterium sprawności jest jednak ustalone odgórnie. W pełni obliczalną regułę zmiany wag synaps zaproponował jako pierwszy Donald Hebb.[53]

W celu uniknięcia ograniczeń, związanych z modelami komputacyjnymi w wyjaśnianiu zmian wag połączeń synaptycznych, należy, zdaniem Penrose'a, odwołać się do zjawiska kwantowej koherencji:

> O koherencji mówimy, gdy duża liczba cząstek tworzy jeden stan kwantowy, który nie ulega splątaniu z otoczeniem. (Termin „koherencja" oznacza, że oscylacje fali w różnych punktach są zsynchronizowane i następują w zgodnym rytmie. W przypadku kwantowej koherencji chodzi o oscylacje funkcji falowej; koherencja oznacza wtedy, że mamy do czynienia z jednym stanem kwantowym.)[54]

Koherentne stany kwantowe obserwuje się w zjawiskach nadciekłości i nadprzewodnictwa, co zostało już powyżej omówione. Penrose wyklucza jednak rolę zjawisk kwantowych w samym zachowywaniu się impulsów nerwowych, gdyż sygnały te są według niego częścią makroświata, a nie mikroświata, i utrzymanie się superpozycji neuronów pobudzonych i niepobudzonych nie wydaje się prawdopodobne z uwagi na redukcję wektora funkcji falowej następującą po wysłaniu impulsu przez neuron.[55] Roger Penrose postuluje zatem znalezienie innej „sceny" dla koherentnych zjawisk kwantowych, które wpływać mają wedle niego na zmiany wag połączeń synaptycznych.

Jeśli uzna się, że jedynym czynnikiem, za pomocą którego wyjaśnić można zachowania zwierząt, jest działanie układu nerwowego, pojawia się oczywisty problem w przypadku organizmów jednokomórkowych. Penrose twierdzi, że istnieć musi inny układ niż układ nerwowy, wpływający na zachowanie jednokomórkowców. Układem takim według niego jest cytoszkielet, który

> (…) Dla pojedynczej komórki nerwowej (…) gra rolę szkieletu, kończyn, układu mięśniowego, krwionośnego i nerwowego.[56]

---

[51] R. Penrose, *Cienie*..., s. 434.
[52] Tamże.
[53] Tamże.
[54] Tamże, s. 431-432.
[55] Tamże, s. 436.
[56] Tamże, s. 439.

Powołując się na prace wspominanego już wcześniej Hameroffa,[57] Penrose twierdzi, iż każda komórka nerwowa budująca mózg wyposażona jest w swego rodzaju własny „układ nerwowy". Cytoszkielet zbudowany jest z trzech elementów: mikrotubuli, aktyny oraz filamentów pośrednich. Mikrotubule są polimerami białkowymi, złożonymi z cząsteczek tubuliny, która występuje w dwóch podstawowych konformacjach przestrzennych jako α-tubulina i β-tubulina. Zdaniem Penrose'a:

> Pewne dane wskazują, że dwie struktury przestrzenne odpowiadają dwóm stanom polaryzacji elektrycznej dimeru, związanym z dwoma możliwymi położeniami elektronu w wiązaniu między α-tubuliną i β-tubuliną.[58]

Mikrotubule budują włókna o przekroju poprzecznym przypominającym wentylator. Włókna te tworzy dziewięć par lub trójek mikrotubuli na około, a przez środek przebiega przeważnie dodatkowo para mikrotubuli. Najważniejszym elementem cytoszkieletu jest struktura utworzona z centrioli, które to z kolei składają się z dwóch cylindrów zbudowanych z dziewięciu trójek mikrotubuli. Centriola jest zasadniczą częścią centrosomu i uczestniczy w procesie podziałów komórkowych.

Kolejnym, ważnym do rozpatrzenia aspektem, jest przytoczona przez Penrose'a teza Hameroffa, głosząca, że mitrotubule są tzw. automatami komórkowymi, przenoszącymi sygnały w postaci fal polaryzacji elektrycznej dimerów tubuliny.[59] Zmiana struktury przestrzennej tubuliny zależna jest od stanów polaryzacji dimeru, co z kolei związane jest z polaryzacją dimerów sąsiednich. Fakt ten pozwala na przesyłanie różnych informacji wzdłuż mikrotubuli. Penrose twierdzi, iż wraz z ukończeniem rozwoju mózgu neurony nie podlegają dalej procesowi podziału, wskutek czego centriole *de facto* nie spełniają dalej swej roli w procesach mitotycznych.[60] Właściwą funkcją mikrotubuli nie jest więc uczestniczenie w mitozie, ale regulacja siły połączeń synaptycznych pomiędzy komórkami nerwowymi, co z kolei warunkuje nieobliczalność procesów zachodzących w mózgu. Kluczową tezę Penrose'a stanowi stwierdzenie, iż nieobliczalność ta spowodowana jest istnieniem *koherentnych zjawisk kwantowych* w mikrotubulach:

> Można przypuszczać — jak pisze Penrose, że natura postanowiła dobrze wykorzystać puste rurki cytoszkieletu. Być może rurki stanowią dobrą izolację, dzięki której kwantowe procesy zachodzące wewnątrz nie ulegają splątaniu z otoczeniem przez dostatecznie długi czas.[61]

---

[57] S. Hameroff, *Ultimate computing: Biomolecular consciousness and nanotechnology*, North-Holland 1987, Elsevier.

[58] R. Penrose, *Cienie…*, s. 441.

[59] Tamże, s. 446.

[60] W świetle współczesnej wiedzy z zakresu neuronauk, także w rozwiniętym mózgu powstają nowe komórki nerwowe, czego w swojej argumentacji nie uwzględnił Penrose. Fakt ten nie podważa jednak w istotny sposób wartości jego koncepcji. Więcej na temat neurogenezy w dojrzałym mózgu znaleźć można w następujących publikacjach: G. Kempermann, F. Gage, *NOWE komórki nerwowe w dojrzałym mózgu*, przeł. A. Bidziński, „Świat Nauki", nr 7, 1999, s. 24-29 oraz M. Sacharczuk, *Neurogeneza wieku dorosłego*, Warszawa 2005, Wydaw. Lekarskie PZWL.

[61] R. Penrose, *Cienie…*, s. 451.

Dodatkowym czynnikiem powodującym, że mikrotubule są układami dostatecznie izolowanymi, by mogły w nich zachodzić zjawiska kwantowe, jest ich wypełnienie przez wodę bez wolnych jonów (obecność filmu wodnego).[62]

Opierając się w większości na badaniach Hameroffa, Penrose utrzymuje, iż dowodem na związek zjawiska świadomości z aktywnością cytoszkieletu jest odwracalny zanik świadomości przy zastosowaniu znieczulenia ogólnego.[63] Choć anestezja wywoływana jest przez różne pod względem chemicznym związki, takie jak chloroform, eter czy tlenek azotu, to czasowy zanik świadomości nie jest, zdaniem Penrose'a, spowodowany siłami chemicznymi, lecz znacznie słabszymi od nich siłami van der Waalsa, które polegają na przyciąganiu pomiędzy cząsteczkami, mającymi moment dipolowy. Penrose twierdzi, że:

> Zgodnie z sugestią Hameroffa i Watta (1983) czynnikiem wywołującym znieczulenie ogólne są siły van der Waalsa (…) uniemożliwiające zmiany struktury przestrzennej dimerów tubuliny. Gazy powodujące znieczulenie ogólne docierają do pojedynczych neuronów, a wtedy momenty elektryczne ich cząstek (…) oddziałują z dimerami, blokując działanie mikrotubuli.[64]

Z uwagi na fakt, iż wymienione powyżej związki wywołują efekt anestetyczny u jednokomórkowców, pozwala to, zdaniem Penrose'a, oczekiwać, iż działają one także na cytoszkielet. Pomimo że działanie cytoszkieletu jest według niego warunkiem *sine qua non* świadomości, to sama świadomość może być zależna od *wielu innych czynników oprócz czynności cytoszkieletu*,[65] takich jak choćby struktura połączeń między komórkami nerwowymi. Dlatego też trudno spodziewać się przejawów jakiejś prymitywnej formy świadomości u takich organizmów jak na przykład pantofelek.

Powstanie świadomości — procesu mentalnego, do którego opisu zmierza omawiana w niniejszej pracy koncepcja Rogera Penrose'a — warunkowane jest przez niealgorytmiczny proces fizyczny, zachodzący na styku mikroskopowej i makroskopowej struktury mózgu. Zdaniem Penrose'a, takim pomostem pomiędzy obydwoma strukturami jest opisywana powyżej procedura obiektywnej redukcji wektora stanu **OR**.[66] W szczególności, istotną rolę odgrywa tutaj kwantowa koherencja, obejmująca nie tylko mikrotubule w jednym cytoszkielecie, ale również poszczególne komórki nerwowe. Koherencja ta ogarniać musi swoim zasięgiem dużą cześć mózgu, wskazując w ten sposób na nielokalny charakter świadomości. Powstały w ten sposób globalny stan kwantowy steruje obliczeniami, zachodzącymi wzdłuż mikrotubuli, regulując w ten sposób siły połączeń synaptycznych.

---

[62] Tamże, s. 451.
[63] Tamże, s. 453.
[64] Tamże.
[65] Tamże.
[66] Tamże, s. 457.

POD OBSTRZAŁEM

Poglądy Rogera Penrose'a na strukturę oraz funkcjonowanie ludzkiego umysłu, które zamieścił w *Nowym umyśle cesarza*, a później rozwinął (jak sam twierdzi *ad nauseam*) w *Cieniach umysłu*, spowodowały znaczącą lawinę polemik ze strony wielu światowej klasy ekspertów, którym bliskie były zagadnienia uwikłane w stawiane przez niego tezy. Spora część tej krytyki znalazła swoje miejsce na łamach elektronicznego interdyscyplinarnego czasopisma *Psyche*, które poświęcone jest szeroko rozumianym badaniom nad świadomością.[67] Do dyskusji włączyli się światowej sławy specjaliści z wielu dziedzin nauki, które w swoich dociekaniach wykorzystał Penrose.

Niniejsze omówienie hipotezy Penrose'a opierać się będzie z jednej strony na argumentach przytoczonych w ramach wspomnianej krytyki oraz stosownych ripost ze strony Penrose'a, a także na wynikach nowszych badań, ze szczególnym uwzględnieniem jego wspólnych osiągnięć ze Stuartem Hameroffem. Bazując jedynie na argumentach w obszarze nauk ścisłych, wyróżnić należy trzy obszary krytyki: (1) zagadnienia matematyczno-logiczne, (2) zagadnienia fizyczne oraz (3) zagadnienia biologiczne z uwzględnieniem kogniwistyki i *neuroscience*. Do tych trzech grup dodać trzeba będzie jednak uwagi natury pojęciowej oraz *stricte* filozoficznej, ponieważ, o ile Penrose „nosi dużo częściej swój kapelusz fizyczny niż filozoficzny", o tyle jego ingerencje w sferę filozofii nie pozostają bez wpływu na całościowy walor jego hipotez.

Największą część polemiki z prezentowaną przez Rogera Penrose'a koncepcją umysłu stanowi wymieniony powyżej jako pierwszy obszar zagadnień matematyczno-logicznych. Przeciwnicy koncepcji Penrose'a zmierzają bowiem do wykazania, iż teza o niealgorytmiczności ludzkiego myślenia nie wynika z przedstawionych przez Penrose'a argumentów na bazie twierdzenia Gödla. Innymi słowy, w wątpliwość podane zostaje twierdzenie, iż — jak pisze Frank Wilczek:

> sposób, w jaki człowiek dochodzi do poznania prawdy, posiada moc dowodową, ale nie może być ujęty w żaden sformalizowany system dowodzenia.[68]

Wilczek słusznie zauważa, iż argument Penrose'a jest *de facto* kontynuacją poglądów oksfordzkiego filozofa Johna Lucasa w tej kwestii, który w wyniku ostrej krytyki nie zyskał powszechnej akceptacji.[69] Trudno jednak pominąć fakt, iż pomimo swej wręcz zaprogramowanej napastliwości, dyskusja na łamach wspomnianego już powyżej czasopisma *Psyche* dostarcza wielu pogłębionych refleksji, pozwalających dokładnie prześledzić, jak Penrose uzasadnia swoje poglądy.

---

[67] *Psyche. An interdisciplinary journal of research on consciousness*, <http://journalpsyche.org>.

[68] F. Wilczek, *A Call for A New Physics. A review of R. Penrose's Shadows of the Mind*, „Science", 266 (1994), 1737.

[69] J. Lucas, *Minds, Machines and Gödel*, „Philosophy", XXXVI, 1961, s. 112-127.

Szczególnie warta przywołania jest polemika autorstwa Davida Johna Chalmersa, gdyż, jak nadmienia w swojej ripoście Penrose,[70] jedynie ona dociera do sedna jego argumentacji. Chalmers zauważa, iż w *Cieniach umysłu* Penrose formułuje *de facto* dwa argumenty przeciwko sztucznej inteligencji. Nieprawdziwość pierwszego argumentu, zakładającego, iż umysł jest systemem formalnym i rozstrzyga o sensowności tego systemu, wynika z faktu, że Penrose operuje zbyt uproszczonym pojęciem takiego systemu jako prostego zbioru aksjomatów i reguł. W rzeczywistości proponenci sztucznej inteligencji naturalnie przyjmują, iż ogromna złożoność systemów, zdolnych do symulacji czynności mózgu takich, jak świadomość, praktycznie pozostawia rozstrzygnięcie o ich sensowności poza zasięgiem metod dostępnych człowiekowi.[71] Natomiast drugi argument Penrose'a, zobrazowany poprzez rozmowę robota z matematykiem, może, w opinii Chalmersa,[72] stanowić pewne wyzwanie dla hipotezy sztucznej inteligencji. W oparciu o swoje rozumowanie Penrose twierdzi, iż założenie o algorytmizacji ludzkiego myślenia prowadzi, zgodnie z twierdzeniem Gödla, do sprzeczności.[73] Według analizy Chalmersa, problem ten nie leży jednak w algorytmizacji, lecz w niezdolności ludzkiego umysłu do przekonania się o sensowności swojego własnego rozumowania. Jak stwierdza Chalmers:

> najprawdopodobniej jesteśmy sensowni, aczkolwiek nie możemy się w sposób niepodważalny o tej sensowności przekonać.[74]

W ostatecznym rozrachunku, Chalmers zarzuca Penrose'owi, iż jego rozważania pozwalają jedynie wyjaśnić, jak powstaje świadomość, a nie czym ona w swojej istocie jest. Z tego powodu, koncepcja Penrose'a nie zasługuje na miano *missing science of consciousness*.

W obszarze dyskusji matematyczno-logicznych warto następnie zwrócić uwagę na krytyczny głos Salomona Fefermana.[75] Feferman podkreśla na wstępie, iż, o ile gödlowskie twierdzenia o niezupełności stawiają istotne pytania co do natury wiedzy matematycznej, o tyle nawet pomimo zdecydowanej sympatii dla niekomputacyjnych modeli mózgu, przedstawiona przez Penrose'a argumentacja jest dla niego nieprzekonująca. Feferman stwierdza, iż

> Celem Penrose'a jest zamiana jednego „nic więcej niż" na drugie: w miejsce stwierdzenia, iż świadomy mózg jest niczym więcej niż działaniem komputera, proponuje tezę, iż świadomy mózg nie jest niczym więcej niż przejawem subatomowej fizyki.[76]

---

[70] R. Penrose, *Beyond the Doubting of a Shadow: A Reply to Commentaries on Shadows of the Mind*, „Psyche" 2 (23) January 1996.
[71] D. J. Chalmers, *Minds, Machines and Mathematics*, „Psyche", 2.16.
[72] Tamże, 3.1.
[73] R. Penrose, *Cienie*... s. 111.
[74] D. J. Chalmers, *Minds*…, 3.14.
[75] S. Feferman, *Penrose's Gödelian Argument*, „Psyche", 2 (7), May 1995.
[76] Tamże, 1.3.

W warstwie czysto formalnej Feferman wypunktowuje wiele logicznych nieścisłości w sposobie, w jaki Penrose wykorzystuje twierdzenie Gödla do wykazania niealgorytmiczności ludzkiego myślenia. Nieścisłości te jednak, zdaniem Fefermana, nie przesądzają jednoznacznie o wadliwości argumentacji Penrose'a. Stanowisko Fefermana wyłania się najpełniej w następującym sformułowaniu:

> Myśl matematyczna nie jest mechaniczna w momencie, w którym powstaje. Zgadzam się z Penrosem, że w tym względzie rozumienie posiada istotną funkcję, i to jest właściwy aspekt matematycznego myślenia, niebędący udziałem komputerów. Poza tym jednak, jego całkowity wysiłek zmierza do uzasadnienia tego przekonania poprzez wykazanie, że w rezultacie twierdzenia Gödla myśl matematyczna nie może być nawet reprezentowana w kategoriach mechanicznych. Moim zdaniem, zamiast podbudowania tego przekonania, wysiłek ten stawia więcej pytań, niż daje odpowiedzi i dlatego jest dialektyką bez wyjścia.[77]

W opinii Fefermana, którą wyraża słowami George'a Boolosa, stwierdzenie, iż umysł może być reprezentowany przez maszynę Turinga (czyli jego działanie jest algorytmizowalne), tkwi nierozwiązana kwestia sposobu, w jaki reprezentacja miałaby się dokonać.[78] Innymi słowy, stwierdzenie braku reprezentowalności umysłu *modo mechanico* wcale nie musi implikować, iż rozumienie jest procesem wykraczającym poza moc maszyn obliczeniowych.

Podsumowując obecnie obszar matematyczno-logicznych kontrowersji wokół gödlowskiego argumentu Penrose'a, warto zajrzeć do chyba najbardziej wyważonej krytyki Daryla McCullougha,[79] z której jasno wynika, iż sformułowanie tego argumentu w postaci sprzeczności prowadzi do wniosku, że jeśli założy się algorytmiczność ludzkiego myślenia, to albo myślenie to jest bezsensowne, albo człowiek nie ma wglądu w to, czym jego rozumowanie jest i czy jest sensowne. Penrose przyjmuje opcję, iż jeśli byłoby pewne, że jakiś algorytm opisuje funkcjonowanie mózgu, to towarzyszyłaby temu też pewność o sensowności tego algorytmu. Natomiast zdaniem McCullougha, gödlowski argument Penrose'a nie stwierdza niealgorytmiczności ludzkiego myślenia, ale tylko sam fakt, iż nie ma przekonujących dowodów na to, że tak jest.[80] Hipotetycznie, umysł mógłby być równie dobrze opisywany algorytmem o takim stopniu złożoności, który uniemożliwia stwierdzenie jego sensowności. Na odrzucaną przez Penrose'a możliwość istnienia zbyt skomplikowanego dla nas algorytmu, który byłby odpowiedzialny za *matematyczny wgląd*, zwraca uwagę w swojej pracy także Stanisław Krajewski, który twierdzi, że autor *Nowego umysłu cesarza* popełnia błędy, gdy korzysta z twierdzenia Gödla, by wykazać niealgorytmiczność *rozumienia.*[81] Skoro więc nie można przeprowadzić sensownego rozumowania w tej materii, wskazuje to na istotne ograniczenia w spójności ludzkiego apa-

[77] Tamże, 4.3.
[78] Tamże, 4.7
[79] D. McCullough, *Can Humans Escape Gödel?*, „Psyche", 2 (4), April 1995.
[80] Tamże, 3.4.
[81] S. Krajewski, *Twierdzenie Gödla i jego interpretacje filozoficzne*, Warszawa 2003, Wydawnictwo IFiS PAN.

ratu poznawczego. Innymi słowy, nie ma możliwości rozwiązania tajników ludzkiego rozumowania jedynie przez „intensywniejsze myślenie".

Znacznie mniejszy ciężar dyskusji wokół Rogera Penrose'a koncepcji umysłu spadł na drugi obszar krytyki, dotyczący podstaw fizycznych tej koncepcji. Nie ulega bowiem wątpliwości, iż to Penrose jest w tym obszarze największym specjalistą, przez co praktycznie jedynym głosem, który potrafi się z pełnym zrozumieniem odnieść do jego koncepcji, jest Stephen Hawking. Powściągliwość komentarza Hawkinga zamieszczonego jako ostatni w *Makroświat, mikroświat…*, nie wynika bynajmniej z kontrowersji natury *fizycznej*, ale *filozoficznej*.[82] Na tym etapie warto jednak powrócić do kilku uwag natury fizycznej, które pojawiły się w dyskusji na łamach *Psyche* oraz wielu późniejszych analiz, i przedstawić je w połączeniu z uwagami biologicznymi. Proponowana przez Penrose'a fizyczna teoria świadomości bezpośrednio dotyczy bowiem obiektów biologicznych, takich jak na przykład mikrotubule.

W pierwszym rzędzie istnieje grono dyskutantów, którzy z bardziej pryncypialnych powodów nie akceptują kwantowych modeli umysłu jako takich. Należy do nich przykładowo polski biofizyk Bernard Korzeniewski. Wychodząc z pozycji silnej sztucznej inteligencji, to jest zupełności opisu procesów mentalnych za pomocą odpowiednio złożonych struktur formalnych, odrzuca on praktycznie każdy fizyczny model mózgu. Jego zdaniem, kwantowy model umysłu jest *redukcjonistyczny*, to znaczy nie rozróżnia pomiędzy kwantowymi efektami, odpowiedzialnymi za świadomość, a innymi makroskopowymi zjawiskami kwantowymi, takimi jak na przykład kondensat Bosego-Einsteina.[83] Argumentując, iż poziom fizyczny (a w tym kwantowy) nie gwarantuje dostatecznej złożoności do wytworzenia zjawisk psychicznych, pomija on chyba jednak fakt, iż kwantowy model umysłu Penrose'a nie bazuje na tak prostych bozonach jak fotony czy też bozony W albo Z, ale dotyczy drobin o masie cząsteczkowej ok. 50 000. Implikuje to niewątpliwie znacznie większy stopień złożoności niż w przypadku wymienionych przed chwilą cząstek elementarnych. Co więcej, Korzeniewski zdaje się w swojej krytyce pomijać istotę zjawiska *dekoherencji,* która, jako jeden z najszybszych procesów zachodzących w przyrodzie, może wpływać na wygaszanie nielokalnych stanów kwantowych w mózgu na skutek sprzężenia z chaotycznymi ruchami molekuł otaczających mikrotubule. W opinii Stephena Hawkinga dekoherencja może w ten sposób efektywnie konkurować z procesami obiektywnej redukcji wektora falowego **OR,** odpowiedzialnymi za generowanie świadomości**,** przez co *de facto* staną się one od dekoherencji nierozróżnialne. Gdyby jednak izolacja mikrotubuli od otoczenia była wystarczająca, wówczas mogłoby się równie dobrze okazać, iż procesy **OR** są zbyt wolne, aby zarządzać zjawiskami mentalnymi.[84]

---

[82] R. Penrose z udz. A. Shimony, N. Cartwrtight, S. Hawking, *Makroświat…*, s. 166.
[83] B. Korzeniewski, *Od neuronu…*, s. 136.
[84] R. Penrose z udz. A. Shimony, N. Cartwrtight, S. Hawking, *Makroświat…,* s. 168.

Okazuje się jednak, iż zagadnienie wzajemnej relacji pomiędzy skalami czasowymi dekoherencji i procesów **OR** poddawane jest dalszym skrupulatnym analizom. Przykładowo, na bazie swoich obliczeń Max Tegmark całkowicie odrzuca możliwość udziału spójnych makroskopowych stanów kwantowych w procesach mentalnych ze względu na niewystarczające ekranowanie mikrotubuli od wpływu środowiska i bardzo szybką dekoherencję rzędu 10^{-20} s.[85] Jego zdaniem, klasyczne modele sieci neuronowych są wystarczające w symulacji funkcjonowania ludzkiego umysłu. Modelu kwantowego bronią z kolei Luiz Pinguelli Rosa i Jean Faber, sugerując zastąpienie obiektywnej redukcji **OR** procesem dekoherencji.[86] Z uwagi na rozbieżność stanowisk oraz złożoność opisu makroskopowych zjawisk kwantowych należy w tym momencie przytoczyć cytowanego już wcześniej Seana Lee, który twierdzi, iż nie ma jeszcze wystarczających teoretycznych ani eksperymentalnych przesłanek co do roli, jaką miałyby spełniać procesy kwantowe w generowaniu zjawisk mentalnych. Podobne stanowisko prezentuje także w serii recenzji w *Psyche* Stanley Klein.[87] Zarzucając Penrose'owi błędną interpretację czasów świadomych odpowiedzi na zadawane impulsy, uzyskanych na bazie doświadczeń Benjamina Libeta,[88] Klein konstatuje, iż klasyczne sieci neuronowe ze sprzężeniem zwrotnym mogą powodować wystarczająco bogatą, spójną aktywność w odpowiednio krótkim czasie. Wynika stąd, iż również w reżimie klasycznym można postulować istnienie stanów sieci neuronowych, obejmujących znaczne obszary mózgu, bez odwoływania się do zjawisk kwantowych.

Krytyka fizycznych podstaw prezentowanej przez Rogera Penrose'a koncepcji umysłu sięga również kwestii najbardziej ogólnej, a jednocześnie chyba najbardziej fundamentalnej. Tyczy się to bowiem polemiki z *fizykalizmem* Penrose'a, zakładającym, iż wszystkie procesy mentalne mogą być ostatecznie zredukowane do zjawisk fizycznych. Warto pamiętać, iż założenie to stanowi podstawę relacji, jaka zachodzi pomiędzy światem fizyki i światem umysłu w promowanej przez niego ontologii trzech światów, o której mowa była we wcześniejszej części niniejszej pracy. W swojej

---

[85] M. Tegmark, *The importance of quantum decoherence in brain processes*, „Phys. Rev.", *E 61* (2000) 4194-4206.

[86] L. P. Rosa, J. Faber*, Quantum Models of Mind: Are They Compatible with Environmental Decoherence?*, „Phys. Rev.", E 70 (2004) 031902.

[87] S. Klein, *Is Quantum Mechanics Relevant to Understanding Consciousness?*, „Psyche", 2 (3), April 1995.

[88] B. Libet, *The natural time-factor in perceptron, volition and free will*, „Metaphisique et de Morale", 2, s. 225-272. Uczestnicy eksperymentu podłączeni do odpowiedniej aparatury pomiarowej mieli za zadanie świadomie wyprostować palec w dowolnie wybranej przez siebie chwili. Wynikiem eksperymentu było stwierdzenie, iż potencjał elektryczny w mózgu wzrastał przed wyprostowaniem palca. Niektóre interpretacje eksperymentu głoszą, iż przeczy on istnieniu „wolnej woli", gdyż mierzony potencjał elektryczny wzrastał przed świadomą decyzją ochotnika. Taki wniosek wydaje się jednak zbytnio uproszczony i nie bierze pod uwagę argumentów natury metodologicznej (np. tego, iż badani byli uprzedzeni, co mają robić, a więc byli już nastawieni na wykonanie pewnych czynności, co tłumaczyć może wzrost potencjału elektrycznego przed domniemaną świadomą decyzją).

bezpośredniej polemice z Penrosem, zamieszczonej w *Makroświat, mikroświat…*[89] Nancy Cartwright argumentuje, iż zjawiska mentalne stanowią przedmiot badań biologii i jako struktura emergentna muszą podlegać autonomicznym prawom biologicznym. Istotnie, jej zdaniem, fizyka (a także chemia) często pomaga w dostarczaniu odpowiedzi na pytania natury biologicznej, to jednak bez użycia analiz biologicznych wyjaśnienie funkcjonowania ludzkiego mózgu nie jest możliwe.[90] Cartwright zauważa, iż fizykalistyczna koncepcja mózgu jest efektem redukcjonistycznego roszczenia, powszechnego wśród fizyków, którzy utrzymują, iż to fizyka samodzielnie dostarczy wyjaśnień ostatecznych bez potrzeby odwoływania się do chemii czy biologii. W pewnym sensie opinia Cartwright jest zbieżna z poglądami Korzeniewskiego. Jego zdaniem, w procesach mentalnych uczestniczą trzy realnie istniejące poziomy rzeczywistości: *fizyczny*, *biologiczny* oraz *psychiczny*, różniące się od siebie stopniem złożoności. Poziom fizyczny, o najniższym stopniu złożoności, jest jedynie „materialnym nośnikiem procesów mentalnych" i sam z siebie nie może generować świadomości.[91]

PODSUMOWANIE

Sformułowanie prostej myśli, podsumowującej Rogera Penrose'a koncepcję umysłu z uwzględnieniem pełnej złożoności tła, na bazie którego powstała, nie jest rzeczą prostą. Czyniąc jednak zadość sugestii, zamieszczonej we wprowadzeniu, postaramy się zaakcentować wątek filozoficzny, który naszym zdaniem stanowi swoistą siłę napędową naukowych dokonań Penrose'a — nie tylko w obszarze teorii umysłu, ale także całokształcie jego naukowego dorobku. Pomimo swojej pozornej oczywistości, sugerowana przez Rogera Penrose'a klasyfikacja stanowisk na temat możliwych związków między zjawiskami mentalnymi a obliczeniami kryje w sobie istotną, dodatkową subtelność. Różnica pomiędzy stanowiskami **A** i **B**, to jest silnej i słabej wersji sztucznej inteligencji, nie sprowadza się, jak to ma miejsce w argumencie Searle'a, bynajmniej do pytania o „świadomość komputerów".[92] Dotyczy ona raczej *źródła* świadomości, ponieważ zgodnie ze stanowiskiem **A** źródłem świadomości jest dostatecznie złożony algorytm, natomiast stanowisko **B** wskazuje na fizyczną strukturę mózgu jako na źródło stanów mentalnych. Innymi słowy, choć strukturę tę można przybliżać algorytmicznie, to jest ona czymś więcej niż tylko siecią formalnych powiązań, niezależnych od substancjalnego podłoża (białko, krzem etc.). Na tej bazie funkcjonuje zresztą argument Chińskiego Pokoju Searle'a, który pokazuje, iż nawet bardzo skomplikowany algorytm nie generuje stanów mentalnych. Algorytmy

[89] R. Penrose z udz. A. Shimony, N. Cartwrtight, S. Hawking, *Makroświat…*, s. 163.

[90] Zob. także B. J. Baars, *Can Physics Provide a Theory of Consciousness?*, „Psyche" 2 (8), May 1995.

[91] B. Korzeniewski, *Od neuronu…*, s. 19.

[92] Zob. np. J. Kloch, *Świadomość komputerów…*

mogą więc te stany doskonale symulować, co nie oznacza, iż mają one w rzeczywistości miejsce — i dlatego słynny test Turinga nie jest w takiej sytuacji rozstrzygający.

Warto jednak dalej zauważyć, iż stanowisko **B** implikuje możliwość symulacji układów fizycznych za pomocą algorytmów, co oznacza, iż świat fizyczny jest *algorytmizowalny*, aczkolwiek nie musi być *algorytmiczny*. Odzwierciedla to w pewnym sensie relację, jaka zachodzi między epistemologicznym postulatem o *matematyzowalności* świata a ontologicznym postulatem o jego *matematyczności.*[93] Ze stanowiskiem **B** można zatem związać *zasadę Churcha–Turinga*, która mówi, iż każdy skończony rzeczywisty układ fizyczny może być doskonale symulowany przez uniwersalną modelową maszynę obliczeniową w skończonej liczbie operacji.[94] W tezie tej (a w konsekwencji w stanowisku **B**) dochodzi do ewidentnego połączenia ze sobą dwóch porządków: *fizycznego* oraz *logiczno-matematycznego*. Nie da się ukryć, iż jest to znaczące roszczenie, a jego uzasadnienie w ostatecznym rozrachunku zależeć będzie od przyjętych założeń natury filozoficznej.

W pierwszym rzędzie Roger Penrose opiera się na przyjmowanym powszechnie przez kognitywistów założeniu *fizykalizmu*, to jest stanowisku głoszącemu, że wszystkie zjawiska mentalne sprowadzalne są ostatecznie do procesów na poziomie fizycznym. W omawianym już obrazie trzech światów Penrose'a, będącym jego „filozofią w pigułce", założenie to tkwi w tym, że świat umysłu jako całość wynika ze świata fizycznego. Ważniejsza jednak konkluzja, stanowiąca centrum prezentowanej w niniejszej pracy dyskusji, staje się widoczna w momencie, gdy porówna się tytuły dwóch głównych części *Cieni umysłu*, w której to pozycji Penrose wyczerpująco przedstawia tajniki swojej koncepcji. Pierwszy z nich brzmi: „Dlaczego potrzebujemy nowej fizyki do zrozumienia umysłu?", drugi natomiast „Jakiej nowej fizyki potrzebujemy do zrozumienia umysłu?". O ile drugi nie budzi żadnych wątpliwości, o tyle pierwszy może zastanawiać choćby z tego powodu, iż zawierając słowo „fizyka" odnosi się do treści pierwszej części *Cieni umysłu*, która o fizyce nie mówi praktycznie nic. Jak się okaże, nie jest to efekt przypadkowy, lecz wynika on z czynionych przez Penrose'a fundamentalnych założeń filozoficznych, pozwalających na, ogólnie rzecz biorąc, daleko idące utożsamienie fizyki z matematyką. Wynika to wprost z relacji, jaka, zdaniem Penrose'a, zachodzi pomiędzy światem matematyki a światem fizyki w obrazie trzech światów.

Spoglądając więc na różnicę pomiędzy stanowiskami **B** i **C** w ujęciu Penrose'a, nie trudno zauważyć, iż różnica ta jest *de facto* dużo mniejsza niż między stanowiskami **A** i **B**. Utrzymując, że nadal struktura fizyczna mózgu odpowiedzialna jest za powstawanie stanów mentalnych, osłabieniu ulega jedynie warunek jej opisywalności za pomocą metod algorytmicznych. Istotnie, taką postawę przyjmuje Penrose, aczkolwiek w jej kontekście dokonuje dość śmiałych konkluzji. W jego opinii udo-

[93] M. Heller, *Filozofia…*

[94] D. Deutsch, *Quantum theory, the Church-Turing Principle and the universal quantum computer*, „Proceedings of the Royal Society", A 400 (1985), s. 97-117.

wodnienie nierozstrzygalności pewnej klasy zagadnień matematycznych przy użyciu maszyny Turinga na bazie twierdzenia Gödla orzeka o niealgorytmicznym charakterze procesów mentalnych i świadomości. Aby jednak taką tezę wysunąć, należy uzasadnić, w jaki sposób czysto logiczno-matematyczny wniosek może mieć konsekwencje fizyczne. Jest to możliwe w kontekście uprawianego przez Penrose'a radykalnego *platonizmu matematycznego*, suponującego obiektywne istnienie świata idei matematycznych, z których wyłania się świat fizyki. Innymi słowy, pojawia się tutaj ontologiczna teza o matematyczności świata fizyki, u podłoża którego leżą struktury matematyczne. Choć Penrose ostatecznie odcina się od identyfikowania rzeczywistości fizycznej z platońskim światem matematyki,[95] to, przykładowo, ścisły związek między formalizmami interpretacji mechaniki kwantowej a ontologiami mikroświata wskazuje, iż granica jest tutaj bardzo płynna.[96] Zmiana natury tych struktur, do których z pewnością należy kwestia algorytmiczności, może więc rzutować na zmianę praw, rządzących rzeczywistością fizyczną. W konsekwencji, matematyczno-logiczne uzasadnienie niealgorytmiczności myślenia uprawnia, jak wskazuje tytuł pierwszej części *Cieni umysłu*, do postulowania nowej, niealgorytmicznej fizyki — *missing science of consciousness*, której czysto techniczne szczegóły rozwija część druga.

Platońskie inspiracje Penrose'a odgrywają również istotną rolę w uzasadnieniu relacji pomiędzy światem umysłu a światem matematyki. Pokazując w swoich wywodach, iż myślenie matematyczne posiada charakter niealgorytmiczny, Penrose łączy ten sposób poznawania prawd matematycznych z bezpośrednim wglądem w platoński świat idei matematycznych. W filozofii Platona taki wgląd stanowi najdoskonalszą formę poznania, gdzie prawda osiągana jest przez bezpośredni (intuicyjny) kontakt umysłu z przedmiotem postrzeganym. Forma ta nosi nazwę νόησις. W odniesieniu do matematyki, Penrose opisuje to w następujący sposób:

> Wyobrażam sobie, że ilekroć umysł postrzega matematyczne pojęcie, styka się z platońskim światem idei (…). Gdy ktoś „widzi" prawdę matematyczną, jego świadomość dociera do świata idei i nawiązuje z nim bezpośredni kontakt — świat ten staje się dlań dostępny za pośrednictwem intelektu. Opisałem to „widzenie" w związku z twierdzeniem Gödla, lecz stanowi ono istotę rozumienia wszystkich prawd matematycznych. Rozmowa między matematykami jest możliwa, ponieważ obaj mają bezpośredni dostęp do prawdy; świadomość każdego z nich może bezpośrednio postrzegać matematyczne prawdy dzięki temu procesowi „widzenia".[97]

Nietrudno się przekonać, iż postulowanie (bo trudno to chyba inaczej określić) bezpośredniego wglądu w niezależnie istniejący świat prawd matematycznych stanowi swoiste „domknięcie" schematu trzech światów Penrose'a. Przytoczona przez niego argumentacja istnienia takiego wglądu na bazie algorytmicznej nierozstrzygalności problemów matematycznych jest w gruncie rzeczy czysto fenomenologiczna i psychologiczna, gdyż odwołuje się do jego osobistego doświadczenia odkrywania

---

[95] R. Penrose, *Droga…*, s. 989-990.
[96] W. Grygiel, *Interpretacje*…, s. 59-72.
[97] R. Penrose, *Nowy umysł…*, s. 469.

prawd matematycznych, niekoniecznie podzielanego przez innych.[98] Istotną rolę w tym uzasadnieniu odgrywa także kwestia prostoty prawdy matematycznej, a także analogii z poznawczym dostępem do takich pojęć jak dobro czy piękno.[99] Choć istnienie intuicyjnego wglądu w świat idei matematycznych jest, obok matematyczności rzeczywistości fizycznej, kolejnym silnym założeniem *filozoficznym*, jakie czyni Penrose, to jednak trudno chyba dzisiaj wskazać równie odważną próbę dostarczenia globalnej ontologii przyrody, określającej tak szczegółowo, jakie miejsce zajmuje w tej ontologii ludzki umysł i świadomość. W końcu sam początek racjonalnej refleksji nad rzeczywistością fizyczną w starożytnej Grecji oparty był na niezwykle mglistych założeniach w porównaniu do precyzji współczesnego języka nauki, a wcale nie okazał się fałszywą ścieżką myślowego wysiłku człowieka. Co więcej, jak już wielokrotnie wspominano, sam Penrose nie traktuje swoich propozycji jako gotowych teorii naukowych, a jedynie jako *tajemnice*, wciąż oczekujące na swoje precyzyjne rozwiązanie w ramach nowej teorii fizycznej. Można również odnieść wrażenie, iż kognitywistyczny aspekt tych propozycji traktowany jest przez Penrose'a jako najbardziej hipotetyczny. W *Drodze do rzeczywistości*, która niewątpliwie stanowi jego naukowe *credo*, koncepcji umysłu poświęcony jest zaledwie (sic!) kilkustronicowy podrozdział z zaznaczeniem jednak, iż, pomimo ostrej krytyki, z jaką się spotkała, jej rola w przyszłości okaże się znacząca.[100] Nie da się ukryć, iż Penrose postawił w tym miejscu na te aspekty swoich rozstrzygnięć, które oparte są na solidnym fundamencie matematycznym. O ile może to sprawiać wrażenie, że Penrose pozbywa się w ten sposób swojego „filozoficznego kapelusza", to jednak daje on jednoznacznie do zrozumienia, iż sformułowanie nowej, zunifikowanej teorii fizycznej, będzie wymagało odwołania się do wartości estetycznych:

> W takim razie dalszy postęp w kierunku pełniejszego zrozumienia fizyki — jeśli nie będzie precyzyjnie kierowany wynikami eksperymentów — musi coraz bardziej polegać na naszej zdolności postrzegania fizycznego znaczenia i głębi matematyki, na wytropieniu właściwych koncepcji i na wyczuleniu na estetyczne walory matematycznej konstrukcji.[101]

Mimo iż trzy światy Penrose'a wydają się posiadać sporą ontologiczną autonomię, stanowią one części jednej, złożonej rzeczywistości. Możliwe jest rozumienie tego schematu w kontekście klasycznego dualizmu podmiot poznający — przedmiot poznawany. Dualizm ten wydaje się dodatkowo wzmocniony przez fakt, iż, jak stwierdza sam Penrose, w jego koncepcji to świadomość stanowi skutek redukcji wektora falowego, a nie — tak, jak sugerowała to interpretacja kopenhaska mechaniki kwantowej — redukcja wektora falowego dokonuje się na skutek interwencji świadomego obserwatora. Wnioskowane przez twórców wczesnej mechaniki kwantowej zacieranie się różnic pomiędzy podmiotem a przedmiotem powraca więc tutaj

---

[98] K. Śleziński, *Elementy platonizmu u Rogera Penrose'a*, Kraków 1998, PAT, 1998, s. 201.
[99] R. Penrose, *Droga…*, s. 989.
[100] Tamże, s. 990-993.
[101] Tamże, s. 987.

do swojej dawniejszej ostrości. Ciekawe jednak, iż Penrose zupełnie pomija kwestię poznawalności świata fizycznego, czyli relacji odwrotnej do kierunku świat fizyki — świat umysłu. Czyżby wgląd w struktury matematyczne był wystarczający do poznawczego spenetrowania świata fizyki? Istotnie, z punktu widzenia platońskiej teorii poznania byłoby to uprzywilejowane poznanie niezmiennego świata idei w stosunku do poznawczo mylącego świata zmysłów. To dlaczego na innym miejscu Penrose określa świat platoński mianem „bardziej prymitywnego"?[102] Odpowiedź na rodzące się w tym momencie pytanie o rzeczywisty obraz platonizmu Penrose'a będzie przedmiotem osobnego studium, rozszerzającego wiele cennych obserwacji, poczynionych już przez Krzysztofa Ślezińskiego.

Patrząc zatem całościowo (i chyba „niealgorytmicznie") na globalną ontologię rzeczywistości Penrose'a, trudno pozbyć się wrażenia, iż, niczym średniowieczny myśliciel pochłonięty kontemplacją prawdy, angażuje on w swoje dociekania niemałą dozę właściwej sobie mistyki. Być może w obliczu współczesnego kryzysu eksperymentalnej weryfikowalności teorii jest to znów, wzorem „windy Einsteina", jedyna droga do zauważenia idei, łączącej w sobie tak koncepcyjnie rozbieżne matematyczne formalizmy ogólnej teorii względności i mechaniki kwantowej w spójną teorię kwantowej grawitacji. Ideą mogącą odegrać znaczącą rolę w zrozumieniu natury czasoprzestrzeni może być rozwijana przez Penrose'a teoria twistorów oparta o „magię liczb zespolonych".[103] I nikt się chyba specjalnie nie pogniewa, jeśli nawet silne założenia filozoficzne zaowocują kolejnym milowym krokiem fizyki.

[102] Tamże, s. 990.
[103] Tamże, s. 925.