1.Rozumienie i algorytmy

Jeśli chodzi o tematy zwykle grupowane pod nagłówkiem „AI" (sztuczna inteligencja), perspektywy dwóch głównych mówców, fizyka matematycznego Rogera Penrose'a i filozofa Emanuele Severino, są oczywiście zupełnie odmienne. Niemniej jednak, jak czytelnik wkrótce odkryje, obaj są zgodni, że nie mamy jeszcze „inteligentnych urządzeń", a także, że jeśli będziemy podążać za wizją tzw. Silnej AI (obecnie wciąż mainstreamowej), nigdy nie będziemy w stanie do budowy takich urządzeń. W swoim głównym wystąpieniu Penrose skupia się na relacjach pomiędzy słowami „inteligencja", „rozumienie" i „świadomość". „Będąc matematykiem" – dzieci – „powiązania między słowami są dla mnie ważniejsze niż ich „prawdziwe" znaczenie". Innymi słowy, relacje między pojęciami są bardziej pouczające niż definicje substancjalne. Dlatego Penrose wychodzi od poglądu, że słowo „inteligencja", przynajmniej w standardowym użyciu, implikuje „rozumienie", a „rozumienie" wymaga pewnej „świadomości" lub „świadomości". Przeglądając różne, szczegółowo omówione przykłady, Penrose pokazuje, że maszyny, czyli oprogramowanie, którymi dzisiaj dysponujemy, to urządzenia obliczeniowe, być może bardzo wyrafinowane, ale wszystkie zasadniczo oparte na idealnym prototypie Maszyny Turinga. Z drugiej strony „inteligencja" i „rozumienie" wydają się wykazywać właściwości, które wymykają się prostej obliczalności. Inteligencja to coś więcej niż tylko zdolność obliczeniowa. Przykłady zaczerpnięte z szachów, indukcji matematycznej (intrygujące twierdzenie Goodsteina), układania płaszczyzny euklidesowej (poliomino, których nie można wytworzyć za pomocą algorytmów obliczeniowych), wszystkie zbiegają się, aby pokazać, że „zrozumienia nie można osiągnąć za pomocą reguł". Zatem ogólna jakość zrozumienia wydaje się taka, że nie jest to algorytm. Zatem „rozumienie" zdaniem Penrose'a nie jest elementem ani wynikiem (bardzo) skomplikowanego zastosowania zbioru reguł (algorytmu).

2. Mechanika kwantowa i świadomość

Następnie Penrose dokonuje przeglądu dwóch głównych filarów teoretycznych współczesnej fizyki, a mianowicie ogólnej teorii względności i mechaniki kwantowej, i wskazuje, że istnieje tylko jeden specyficzny element współczesnej fizyki, którego nie można odtworzyć na komputerze, ponieważ nie można go obliczyć. A dokładniej jest to proces pomiarowy w mechanice kwantowej, tzw. „załamanie się funkcji falowej". Nie jest to opisane równaniem Schroedingera i nie może zostać zaimplementowane (nawet w zasadzie) żadnym algorytmem obliczeniowym. Jak stwierdził Penrose: „Idea jest taka, że procesu zapadania się nie da się obliczyć. W rzeczywistości jest to coś, co przypomina trochę wolną wolę. Ponieważ zgodnie z obecną fizyką podejmuje własne decyzje. W jakiś sposób decyduje się być tu czy tam." Zatem w całym krajobrazie teorii i zjawisk fizycznych wydają się istnieć tylko dwie rzeczy charakteryzujące się wewnętrzną, niealgorytmiczną, nieprzeliczalną naturą: proces pomiaru w mechanice kwantowej (lub załamanie się funkcji falowej) oraz zjawisko „rozumienia" lub „świadomości" właściwe świadomości (ludzkiej). W swojej twórczości Penrose od wielu lat próbuje połączyć te dwie koncepcje. Zaczynając w szczególności od książki „Nowy umysł cesarza" formułuje teorię, zgodnie z którą pochodzenia niealgorytmicznych procesów „rozumienia", „świadomości" i „świadomości" należy upatrywać w pewnych procesach kwantowych zachodzących w określonych obszarów mózgu. Dzięki jego współpracy z anestezjologiem Stuartem Hameroffem w latach 90. XX wieku w „mikrotubulach zidentyfikowano kilku obiecujących kandydatów na gospodarzy tak zwanych „spontanicznych redukcji orkiestrowych" stanów kwantowych. Mikrotubule to bardzo małe struktury subkomórkowe, zlokalizowane wewnątrz neuronów, w aksonach i dendrytach. W tych strukturach spójność kwantowa może być utrzymana wystarczająco długo, aby kwantowe załamanie funkcji falowej mogło bezpośrednio wpłynąć na „elementy" lub „atomy" świadomości kwantowej, zwanej „protoświadomością". Według Hameroffa i Penrose'a mogą to być elementy, z których zbudowana jest

świadomość. Oczywiście w tych cegiełkach nie ma jeszcze świadomego celu, jeszcze celu, a tym bardziej znaczenia, ale z nich mogą wyłonić się struktury zawierające świadome zachowanie.

3. Protoświadomość

Bezpośrednią konsekwencją tej teorii jest to, że świadomość nie jest cechą czysto ludzką, lecz powinna pojawiać się wszędzie tam, gdzie obecne są struktury takie jak mikrotubule, a więc na przykład u zwierząt takich jak małpy człekokształtne, delfiny, psy, koty czy myszy. Ta wizja, również pełna implikacji etycznych, była tym, co wywołało dyskusję z Severino. Co więcej, Penrose pokrótce przedstawił także możliwość „zbudowania" świadomego, a zatem inteligentnego urządzenia. Jego zdaniem można tego oczywiście dokonać jedynie poprzez agregację elementów protoświadomości i zapewnienie im odpowiedniego środowiska, w którym może nastąpić „zaaranżowana obiektywna redukcja" funkcji falowej. Całkiem oczywiste jest, zdaniem Penrose'a, że dzisiejsze komputery, a najprawdopodobniej także te jutra, opierają się na możliwościach czysto algorytmicznych, są i pozostaną „nieświadome", a zatem pozbawione prawdziwego zrozumienia i inteligencji. W szczególności nie ma niebezpieczeństwa, że pewnego dnia „inteligentne maszyny" będą w stanie przejąć władzę nad światem i zagrozić rasie ludzkiej lub ją zniszczyć. Powszechnie wiadomo, że Stephen Hawking miał w tej kwestii odmienne zdanie.

4. Wolna wola i osobliwości

W związku z tematem wolnej woli i w świetle powyższych idei możemy podjąć dalsze rozważania. Jak powszechnie wiadomo, zawodowi ogólni relatywiści obawiają się osobliwości, które mogą pojawić się w ogólnej teorii względności. Roger Penrose zdobył Nagrodę Nobla w dziedzinie fizyki w 2020 r. tylko dlatego, że w swojej przełomowej pracy z 1965 r. udowodnił twierdzenie, które stwierdzało, że w bardzo ogólnych warunkach (w szczególności bez konieczności stosowania symetrii sferycznej) osobliwości są ogólną i nieuniknioną przewidywaniem Ogólnej Teorii Względności (zarówno w kierunku przeszłym, czyli osobliwości kosmologicznej Wielkiego Wybuchu, jak i w kierunku przyszłym, czyli osobliwości czarnych dziur). Ale dlaczego fizycy teoretyczni nie lubią osobliwości? Standardowa odpowiedź jest taka, że w przypadku osobliwości zdolność przewidywania teorii całkowicie zawodzi. Z osobliwości dosłownie wszystko może wyjść (lub wejść) w całkowicie nieprzewidywalny sposób, ponieważ prawa fizyczne z definicji rozbijają się w osobliwości. W rzeczywistości, aby chronić obserwowalny wszechświat przed takim potworem, Penrose dawno temu zaproponował hipotezę kosmicznej cenzury, zgodnie z którą osobliwości są zawsze owinięte i ukryte za horyzontem zdarzeń. Dlatego nie mogą wpływać na wszechświat zewnętrzny. Jednakże, zgodnie z powyższymi rozważaniami Penrose'a, istnieje również inne zjawisko, które przedstawia punkty „osobliwości", w których zdolność predykcyjna teorii całkowicie zawodzi: jest to pojedyncze zdarzenie w mechanice kwantowej! Przy każdym załamaniu się funkcji falowej dochodzi do awarii obliczalności, stąd „osobliwości", w których zawodzi zdolność przewidywania teorii kwantowej. Jeśli na przykład weźmiemy pod uwagę emisję pojedynczego fotonu ze wzbudzonego atomu, to cóż, nie da się przewidzieć ani chwili, ani kierunku tej emisji. Analogicznie dzieje się w przypadku rozpadu neutronu: dla pojedynczego zdarzenia nie jest znana ani chwila, ani kierunek emisji pary neutrino-elektron. Teoria kwantowa przewiduje jedynie rozkłady prawdopodobieństwa, które choć doskonale zgadzają się z eksperymentem, z definicji odnoszą się tylko do klas zdarzeń, a nie do pojedynczych zdarzeń. Dlatego „osobliwości", czyli punkty, w których zawodzi przewidywalność, wydają się być obecne wszędzie w teorii kwantowej.

5. Konstruowanie świadomości

Dwa z tematów, które Emanuele Severino wybiera do krytyki, to „miejsce", w którym może znajdować się ludzka inteligencja lub świadomość, oraz możliwość „zbudowania" inteligentnego (lub świadomego) urządzenia. Severino zaczyna od wskazania hipotetyczno-dedukcyjnego charakteru

nauki. Wszystkie nauki, a w szczególności nauki matematyczne, opierają się na postulatach, czyli sądach uznawanych za oczywiste, z których według pewnych (logicznych) reguł wynikają inne twierdzenia. Ale postulaty, a nawet reguły dedukcyjne nie są uważane za prawdy niepodważalne, nawet przez samą Naukę. W szczególności same w sobie są „konwencjami". Lub, używając słowa, któremu Severino przypisuje szerokie znaczenie, są to „wiary". Czyli wyrazy „woli mocy". Według Severino „o wyborze między dwiema różnymi i konkurującymi teoriami ostatecznie decyduje zdolność jednej, a nie drugiej, do przekształcenia świata". Sama nauka uznaje, że nie jest już możliwe budowanie wiedzy, której nikt, ani bogowie, ani ludzie, nie może zaprzeczyć. Dlatego też, zdaniem Severino, nauka nie ma na celu znalezienia niepodważalnej prawdy; dąży raczej do posiadania władzy nad światem. Cel ten jest sprzeczny z tym, co Severino uważa za pierwotny, fundamentalny cel filozofii, cel, dla którego filozofia narodziła się 2500 lat

6. Modelowanie matematyczne

Jednakże pogląd Severina, że starożytni greccy matematycy, a także inni, aż do Galileusza, próbowali dojść do niezaprzeczalnych prawd epistemicznych („znać twierdzenia tak, jak je zna Bóg", jak mówi Galileusz), z pewnością można skrytykować. W rzeczywistości jest dość jasne, zwłaszcza według najnowszych rekonstrukcji historycznych (patrz np. Lucio Russo), że koncepcja „modelu matematycznego" była już dobrze rozwinięta w pracach Euklidesa, Archimedesa i innych. Swobodny wybór postulatów, możliwość „bawienia się" nimi w celu lepszego modelowania danego zjawiska, czy danego zbioru twierdzeń (matematycznych) to działania bardzo jasno określone i skutecznie wykonywane przez starożytnych matematyków, aż do współcześni ojcowie rewolucji naukowej, Kopernik, Galileusz, Newton itd. Właściwie można śmiało stwierdzić, że żaden z nich tak naprawdę nie wierzył w konstruowanie prawd absolutnych; zamiast tego zajmowali się modelami (matematycznymi), które radziły sobie lepiej niż poprzednie (np. znalezieniem fizyki lepszej niż fizyka arystotelesowska lub astronomii lepszej niż astronomia ptolemejska itp.)

7. Manifestacja świata

Następnie Severino wprowadza koncepcję „manifestacji świata", wymiaru, w którym wszystko się dzieje, wymiaru, w którym wszystko się manifestuje, wymiaru, z którego czerpie się wiadomości o każdym pojedynczym fakcie lub rzeczy: „Nauka nie może zrobić żadnego kroku która nie wypływa z przejawienia się świata." Według Severina „objawienie świata" nie jest rzeczą między innymi, ponieważ zawiera w sobie całą przeszłość, teraźniejszość i przyszłość świata, wszystko, co się pojawia . Do idei „manifestacji świata" Severino odwołuje się także w swojej drugiej wypowiedzi, w której zdecydowanie sprzeciwia się poglądowi Penrose'a na temat „miejsca (w mózgu), w którym należy szukać świadomości". Według Severino szukanie miejsca, „w którym przebywa świadomość", oznacza w istocie ponowne rozważenie świadomości jedynie jako określonej części owego „pojawiania się świata", owego „przejawu świata", który w istocie należy rozpatrywać jako pierwotna forma świadomości. Niemniej jednak ta idea „świadomości świata" wydaje się mieć wspólne cechy z wprowadzonymi przez samego Penrose'a „atomami" protoświadomości (wspólnymi dla różnych bytów), a podobieństwa można doszukać się także w koncepcjach rozważanych przez innych autorów tej pracy. książkę, Vitiello, D'Ariano i Faggin, choć z różnymi akcentami i oglądanymi pod różnymi kątami.

8. Produkcja

Istota krytyki Severino dotyczącej idei „sztucznej" inteligencji, czyli „sztucznej" świadomości, skupia się na idei samej „produkcji". Jak zwykle w stanowisku filozoficznym Severina, błąd, główny nihilistyczny błąd zachodniej filozofii (cywilizacji), kryje się w czasowniku „wyprodukować", w idei „konstrukcji", tj. „produkcji" czegoś. Pojęcie produkcji, jak mogłoby się wydawać, nie jest niewinne, ale kryje się w nim

oczywiste, zdroworozsądkowe przekonanie, że rzeczy mogą powstać z niczego i (w razie potrzeby) wrócić, wrócić do wrócić do niczego. Według Severino obserwacja, że rzeczy można tworzyć i niszczyć, przekonanie, że rzeczy oscylują pomiędzy byciem i niebytem, z towarzyszącym przejściem od niebytu do bytu, a co za tym idzie stawaniem się rzeczy, co wszyscy uważamy za najwyższe oczywistość, nie jest przedmiotem obserwowalnym, nie jest treścią obserwacji, ale jest teorią, a mianowicie jedną z wielu możliwych interpretacji rzeczywistości. Za tą upartą wiarą kryje się duch nihilizmu: wiara, że rzeczy są właściwie niczym. Jest oczywiste, że z tego punktu widzenia idea wytworzenia sztucznej inteligencji jest po prostu kolejnym przejawem głębokiej nihilistycznej wizji przenikającej cywilizację zachodnią

9. Sztuczna inteligencja

Ale Severino idzie dalej. Zauważa, że platońska definicja „produkcji" jako „przyczyny, która sprawia, że rzecz przechodzi z niebytu do bytu", przenika całą myśl zachodnią, w filozofii, ekonomii, prawie i naukach matematycznych, i zawsze jest uważana za oczywiste. Co więcej, myśl zachodnia postrzega człowieka jako jedyną istotę zdolną do organizowania środków do realizacji celów. Ale, jak zauważa Severino, jest to również definicja maszyny! Tyle że na razie maszyny nie mają żadnego celu na celowniku: „Człowiek jest tą maszyną, która organizuje środki w celu wytwarzania celów, mając na uwadze właśnie obecność celów, idealną obecność". W ten sposób Severino dochodzi do prowokacyjnego wniosku, że „naturalnego" człowieka uważa się za maszynę, a sam świat postrzega się jako mechanizm, w ramach którego zorganizowane są środki w celu osiągnięcia celów. A zatem, biorąc pod uwagę sposób, w jaki Człowiek został rozumiany na Zachodzie, wówczas Człowiek, a raczej naturalna inteligencja Człowieka, „jest" już inteligencją „sztuczną". W przeciwieństwie do tego Severino potwierdza, że „przejaw świata jako całości – ta pierwotna i podstawowa forma świadomości – nie może być przedmiotem produkcji przynajmniej z tego powodu: ponieważ producent, gdyby miał wytworzyć całość przejawu świata, musiałoby znajdować się poza przejawem świata i dlatego byłoby czymś nieznanym.

10. Debata

Druga część dialogu zawiera interakcję, czyli debatę, pomiędzy Penrose'em i Severino oraz pomiędzy publicznością a mówcami. Wiele pytań dotyczyło kwestii mikrotubul. Jeśli chodzi o ich zdolność do przetrwania śmierci, a tym samym zachowania niektórych wspomnień z poprzedniego życia, Penrose jest zdecydowanie sceptyczny: „Uważam, że mikrotubule nie przetrwają śmierci bardziej niż neurony". Drugi zestaw pytań dotyczył testów świadomości, a w szczególności sposobów badania wpływu środków znieczulających na świadomość. Panuje ogólna opinia, że mikrotubule są bardzo zaangażowane w działanie środków znieczulających. Penrose wprowadza tu także rolę móżdżku, który steruje wszystkimi „automatycznymi" działaniami człowieka i którego działanie wydaje się całkowicie nieświadome, w przeciwieństwie do roli mózgu, gdzie świadome działania wydają się odgrywać wyraźną rolę . Na potwierdzenie swoich poglądów Penrose wskazuje, że mikrotubule występują obficie w komórkach piramidalnych, które z kolei występują obficie w mózgu, podczas gdy nie występują one w móżdżku. Interesujące jest to, że świadomość wydaje się pojawiać w komórkach piramidalnych. Inne testy świadomego zrozumienia przeprowadzane są z konkretnymi problemami szachowymi, przeznaczonymi dla ludzi i komputerów. Widać tu różnicę pomiędzy „rozumieniem", np. tego, co robi bariera pionków, a czysto „mechanicznym" obliczeniem, które wykonuje komputer. Testy te wyraźnie ilustrują różnicę pomiędzy świadomym myśleniem (lub świadomym rozumieniem) a zwykłymi obliczeniami.

11. Świadomość u zwierząt

Wręcz przeciwnie, zdaniem Penrose'a pojęcie kreatywności jest mylące i dwuznaczne: kreatywność nie jest dobrym testem świadomości. Zrozumienie to coś, w czym można dostrzec różnicę pomiędzy świadomym i nieświadomym działaniem. Zamiast tego bardzo trudno dostrzec różnicę między kreatywnością a po prostu przypadkową produkcją czegoś (co jest uważane za odmienne od tego, co zostało wcześniej „stworzone"). Zatem kreatywność, w przeciwieństwie do zrozumienia, jest niezwykle trudna do przetestowania i trudno ją przetestować obiektywnie w tej kwestii. Wreszcie Penrose popiera pogląd, że zjawisko świadomości występuje także u zwierząt. Jest to spójne z jego wizją mikrotubul jako areny, z której wywodzą się zdarzenia protoświadome, a w końcu sama świadomość. Mikrotubule są bardzo obecne u wielu „wyższych" zwierząt i dlatego powinna istnieć świadomość, w szczególności u psów, krów, słoni, małp, goryli, delfinów, myszy itp. Z tego punktu widzenia można wyciągnąć wnioski etyczne, takie jak jako szacunek, jaki jesteśmy winni nie tylko innym ludziom, ale także (wielu) innym stworzeniom.

12. Świadomość i język

Idea zaproponowana przez Penrose'a i Hameroffa (świadomość wyłania się z kumulacyjnego procesu elementarnych aktów „protoświadomości", aż w końcu dotrzemy do cudów ludzkiego umysłu) wymaga niepokojącej obserwacji na temat „silnego" programu AI. Jak powszechnie wiadomo i jak to miało miejsce w przeszłości, program taki zaczyna się od języków (formalnych) i ma na celu odbudowę „inteligencji" poprzez oprogramowanie (tj. programy komputerowe oparte na językach), postępując w pewnym sensie od góry do dołu. Dzieje się to z dość jawnym przekonaniem, że być może od konstrukcji inteligencji uda nam się następnie przejść, podążając tą samą ścieżką, do konstrukcji świadomości. Jednakże, zgodnie z koncepcjami Penrose'a, wydaje się, że Natura jako pierwsza zaczęła konstruować proste formy świadomości, co w rzeczywistości wydaje się być zjawiskiem dość powszechnym, przynajmniej u „wyższych" zwierząt. Dopiero potem Natura buduje języki (nawet skomplikowane); a języki, przynajmniej te zaawansowane, wydają się być istotne tylko dla jednego gatunku – ludzi. Z tego punktu widzenia program Silnej AI wygląda na prawdziwie „sztuczny", w tym sensie, że podąża ścieżką przeciwną do tej, którą podąża naturalna ewolucja. Ludzie próbują budować świadomość zaczynając od języka, podczas gdy Natura zbudowała języki zaczynając od (proto)świadomości! Z drugiej strony, jak powiedział kiedyś D. Hofstadter, być może sztuczną inteligencję należałoby porównać do współczesnego samolotu odrzutowego: osiągającego wysokie osiągi, jeśli chodzi o konkretne zadania, ale generalnie niezdolnego do wykonywania takich rzeczy, jak wróbel, który w metaforze reprezentuje człowieka inteligencja, może to z łatwością zrobić. Odrzutowiec może przelecieć z Londynu do Mediolanu w godzinę, co jest zadaniem niemożliwym do wykonania dla wróbla czy gołębia. Ale spróbuj wylądować odrzutowcem na rynsztoku...

13. Miejsce dla świadomości?

Kontrowersja Severino skierowana do Penrose'a dobrze ilustruje dystans i różnice między ich podejściami. Severino otwarcie krytykuje Penrose'a za zignorowanie jego słów o „ujawnieniu się świata jako pierwotnej formy świadomości", choć koncepcja ta jest obecna, zdaniem Severino, u wielu ważnych przedstawicieli współczesnej kultury, takich jak Kartezjusz, Kant, Brouwer. Kiedy Penrose wielokrotnie szuka „miejsca, w którym przebywa świadomość", po prostu pokazuje, zdaniem Severino, że nadal wyobraża sobie świadomość jako rzecz pośród rzeczy, a mianowicie jedynie jako część w obrębie tego „przejawu świata", który jest w rzeczywistości pierwotnym formę świadomości. Ostatnim tematem, w którym rozbieżność jest szczególnie wyraźna, jest praktyczny charakter nauki, co według Severino oznacza to następująco: pojęciowe sformułowania wiedzy naukowej pozwalają na władzę nad światem przewyższającą inne pojęciowe artykulacje, takie jak pojęciowe sformułowanie sojuszu z świętość, czyli modlitwa. Zatem pojęciowa artykulacja współczesnej wiedzy naukowej jest imponująca! To właśnie pozwala dziś na największą władzę nad światem. Ale władza to jedno, a prawda to drugie.

Severino powraca do swojego początkowego argumentu dotyczącego technologicznej mocy teorii naukowych. Jest przekonany, że ostatecznie teorię naukową ocenia się na podstawie jej technicznej zdolności do przekształcania świata, a nie zdolności do prawdziwego jej przedstawienia lub skutecznego wyjaśnienia. Temat intersubiektywności w nauce przywołuje tu Severino. Zdaniem Poppera, aby istniała władza, potrzebne jest intersubiektywne rozpoznanie tej władzy, a to oznacza, że inni muszą jasno dostrzec przemianę świata.

14. Nauka i technologia

Na tę ostatnią kwestię Penrose odpowiada stanowczo (i moim zdaniem słusznie), potwierdzając tradycyjny rozdział nauki i technologii. W tym miejscu do dyskusji wchodzą głębokie kwestie etyczne. Jak mówi Penrose: „Mój sposób patrzenia na naukę polega na próbie ustalenia prawdy o świecie. Zatem nie ma tu żadnego problemu moralnego. To znaczy, kwestia moralna to osobna kwestia. Nauka próbuje zrozumieć sposób, w jaki działa świat, jak to działa. Następnie jest technologia, która pracuje nad tym, jak wykorzystać wiedzę naukową. Technologia ma ścisły związek z nauką, ale nie jest nauką. Technologia ma ogromny i ciągły wpływ na to, co ludzie robią w codziennym życiu. A to oczywiście rodzi kwestie moralne. Zatem według Penrose'a wykorzystanie technologii, a w szczególności dobre lub złe wykorzystanie nauki, jest głęboko splecione z moralnością. Penrose jasno to stwierdza: „Kiedy zajmuję się nauką, próbuję zrozumieć sposób, w jaki działa świat i nie szukam władzy". Nauka nie próbuje kontrolować świata. Taki jest cel technologii. Technologia i moralność to obszary wyraźnie oddzielone od nauki, choć od niej zależne; kiedy nauka się zmienia, te dwa pozostałe obszary muszą okazać jej szacunek i zobaczyć, co im powie. Na koniec Penrose ponownie wskazuje, że koncepcje jego i Hameroffa skłaniałyby się do uznania świadomości za zjawisko, które nie ogranicza się do ludzi, ale pojawia się także u innych zwierząt. Oznacza to, że kwestia moralna jest związana ze sposobem, w jaki postępujemy ze zwierzętami, i jest to kolejny przykład tego, jak nauka wpływa na nasze przekonania moralne.

15. Chatboty

Esej Ines Testoni, byłej uczennicy Severina, przedstawia szereg interesujących punktów, z których tutaj mogę omówić tylko kilka. Esej otwiera prezentacja eksperymentu, faktycznie przeprowadzonego w 2017 roku w grupie Facebook Artificial Intelligence Research (FAIR). Dwa programy komputerowe (oprogramowanie) zostały przeszkolone w zakresie rozmów w języku angielskim, a następnie umożliwiono im autonomiczną rozmowę ze sobą, używając języka angielskiego, także w sposób nieludzki. Wydawało się, że oba chatboty stopniowo wymyślają język niedostępny dla ludzi. Nawiązany między nimi dialog można wręcz interpretować jako ukonstytuowanie się autonomicznej świadomości w komputerach. Być może – prowokuje Testoni – systemy sztucznej inteligencji przewidują, a nawet zdają sobie sprawę z powstania „kwantowej maszyny Turinga", której ludzie jeszcze nie wiedzą, jak skonstruować. Ta interpretacja doskonale zgadzałaby się z wizjami Putnama i Chalmersa, którzy utrzymują, że materialna budowa umysłu jest całkowicie nieistotna dla wytwarzania myśli i świadomości. Innymi słowy, właściwości psychiczne są organizacyjnie niezmienne w tym sensie, że wsparcie materialne może się zmieniać. Jeśli jakikolwiek stan psychiczny jest organizacyjnie niezmienny, wówczas teoretycznie możliwe byłoby zastąpienie istoty szarej systemem sztucznej inteligencji po śmierci mózgu. Tę intrygującą perspektywę eksplorowali także D. Hofstadter i D. Dennett w słynnej książce „The Mind's I", a w innym kontekście W. Pfister w pięknym i niepokojącym filmie „Transcendencja".

16. kwantowych maszyn Turinga

Jak wiemy, Roger Penrose głęboko nie zgadza się z tezą o „silnej sztucznej inteligencji": systemy sztucznej inteligencji działają wyłącznie poprzez język formalny, podczas gdy myślenie ludzkie

charakteryzuje się funkcjonowaniem, którego nie można sprowadzić do procesów obliczeniowych. Jeśli rdzeniem świadomości (ludzkiej) jest zbiór nieobliczalnych, niealgorytmicznych procesów zapadnięcia się kwantów zachodzących w mikrotubulach, to oznacza to, że świadomości (ludzkiej) nie da się przedstawić za pomocą żadnej konwencjonalnej maszyny Turinga i że ludzki umysł ma możliwości, których nie mógłby posiadać żaden system sztucznej inteligencji ze względu na nieprzeliczalną fizykę związaną z mechanizmem OrchOR. Jedynym mechanizmem, który przynajmniej w zasadzie może uciec od tego stanu rzeczy i dzięki temu mieć szansę „wykazać" świadomość, jest sama maszyna Turinga oparta na fizyce kwantowej, czyli kwantowa maszyna Turinga. Z powyższego Testoni wysuwa intrygującą tezę: mechanika kwantowa może być wspólnym elementem materii organicznej i nieorganicznej, który powoduje umysł i świadomość. Jeśli tak, nie możemy powiedzieć, że świadomość jest tylko ludzka, tylko ze względu na mechanikę kwantową. Wręcz przeciwnie, świadomość może w zasadzie pojawiać się zarówno w materii organicznej, jak i nieorganicznej właśnie dlatego, że mechanika kwantowa leży u podstaw wszystkich znanych form materii. W swoich wnioskach Testoni wskazuje, że według Severino „świadomość jest fenomenologicznym przejawem wszystkiego, co jest wieczne jak każda materia i jest ze sobą tożsame, a zatem nie można jej sprowadzić do materii. Relacji pomiędzy świadomością i materią (istotą szarą w mózgu lub inną) nie można sprowadzić do ich wzajemnej tożsamości." Jeśli świadomość (przejaw) nie ogranicza się do jamy mózgu, to musimy uznać, że wykracza ona poza wymiar indywidualny i ludzki, a tym samym zdolność człowieka do rozpoznania jej obecności. Wyłania się tu związek pomiędzy koncepcją Severino o świadomości ogólnej jako „manifestacji świata" a koncepcją Penrose'a-Hameroffa o „elementach protoświadomości", która powinna być obecna wszędzie tam, gdzie załamuje się funkcja falowa.

17. błędów popełnianych przez maszyny

Esej Giuseppe Vitiello ilustruje najważniejsze aspekty dyssypatywnego kwantowego (teorii pola) modelu mózgu (i świadomości). Już z tytułu Vitiello przypomina zasadniczą rolę, jaką odgrywa (rozpraszający) chaos w zdolności mózgu do elastycznego reagowania na świat zewnętrzny. Freeman podkreślił tę koncepcję i przypisaną Arystotelesowi uwagę, że „mózg nie jest głupią gwiazdą", obrazowo ukazuje fakt, że mózg w swojej wiecznej trajektorii nigdy nie przechodzi przez ten sam punkt w sposób w pełni przewidywalny. Mózg zachowuje się jak „maszyna popełniająca błędy", czyli jak urządzenie z natury błędne. „Spójność" to centralna koncepcja podejścia Vitiello do odkrywania cudów (ludzkiego) mózgu. Z obserwacji wynika, że aktywność neuronalna w korze nowej wykazuje powstawanie rozszerzonych konfiguracji ruchów oscylacyjnych. Konfiguracje te rozciągają się na obszary o wymiarach liniowych do dwudziestu centymetrów w ludzkim mózgu i mają prawie zerową dyspersję fazową. Obecność pewnego rodzaju „współpracy" na tak dużych obszarach oznacza, że funkcji mózgu nie da się wyjaśnić jedynie znajomością poszczególnych elementarnych komponentów. Aktywność mózgu wymaga wprowadzenia pojęcia „koherencji": powszechnej współpracy ogromnej liczby neuronów w rozległych obszarach mózgu. Naturalnym narzędziem matematycznym do opisu koherencji jest Kwantowa Teoria Pola (QFT), ponieważ jej formalizm okazał się bardzo przydatny w badaniu układów biologicznych w ogóle, a mózgu w szczególności. Formalizm matematyczny opisujący „spójność" nadaje dobrze zdefiniowane znaczenie pojęciu wyłonienia się makroskopowej właściwości z mikroskopijnego procesu dynamicznego. Układ makroskopowy posiada właściwości fizyczne, których nie można znaleźć na poziomie mikroskopowym. Vitiello podkreśla, że użycie QFT zamiast mechaniki kwantowej (QM) do opisu mózgu jest konieczne, ponieważ QFT pozwala na opis różnych faz, w których można znaleźć system. Technicznie rzecz biorąc, dzieje się tak, ponieważ w QFT istnieje nieskończenie wiele jednostkowo nierównoważnych reprezentacji kanonicznych reguł komutacji (CCR), podczas gdy QM, przeciwnie, pozwala na jedynie unitarnie (a zatem fizycznie) równoważne opisy (dla układów o skończonej liczbie stopni swobody). Systemy, które mogą mieć różne fazy fizyczne, takie jak mózg,

muszą zostać opisane za pomocą QFT, co może wyjaśniać wielość ich faz i przejść między nimi, do czego QM nie jest przygotowany.

18. Rozpraszający mózg kwantowy

Kolejną fundamentalną obserwacją Vitiello jest to, że mózg jest systemem otwartym, w pełnej interakcji z otoczeniem. Ta strukturalna otwartość doprowadziła Vitiello i innych do sformułowania rozpraszającego kwantowego modelu mózgu. Kluczowym aspektem modelu jest identyfikacja nabywania „nowej" specyficznej pamięci, z określonym stanem podstawowym (spośród nieskończenie wielu, unitarnie nierównoważnych stanów), zwanym także „próżnią", do której mózg-środowisko system ma dostęp. Zgodnie z kwantowym modelem rozpraszającym mózg, na tym właśnie polega „sekret" pamięci. Ten zbiór stanów pamięci, czyli zbiór próżni, można opisać jako „krajobraz atraktorów". Myśli są w zasadzie pojmowane jako posiadające chaotyczne trajektorie w tym krajobrazie próżni. Każdy akt rozpoznania, skojarzenia z konkretnym wspomnieniem można przedstawić jako podejście do pobliskiego atraktora i w konsekwencji jego uchwycenie. Według Vitiello oznacza to akt wiedzy intuicyjnej, uznanie zbiorowego, spójnego trybu, z natury „nieobliczeniowego" i nieprzekładalnego na logiczne ramy języka. Ostateczny „nieobliczeniowy" charakter tej myśli pojawia się tu ponownie, podobnie jak w poglądach Rogera Penrose'a. Trajektorie w krajobrazie atraktorów, od pamięci do pamięci, są klasycznymi trajektoriami chaotycznymi, chociaż „łączą" stany kwantowe; nie są one okresowe (trajektoria nigdy się nie przecina), a trajektorie, które mają różne warunki początkowe, nigdy się nie przecinają; zamiast tego rozchodzą się (wykładniczo). Tutaj początkowa intuicja Freemana znajduje rygorystyczny wyraz: chaotyczne trajektorie dają początek zdolności mózgu do elastycznego reagowania na świat zewnętrzny i generowania nowych wzorców aktywności, w tym tych, które są odbierane jako świeże pomysły. To wędrowanie jest charakterystyczną cechą aktywności mózgu, myślenia. Dlatego mózg nie jest głupią gwiazdą. Zachowuje się raczej jak „urządzenie błędne", „maszyna popełniająca błędy". Według Vitiello nieobliczeniowa aktywność błędnego urządzenia powinna być cechą charakterystyczną inteligentnego urządzenia: dokładnie tym, co według Penrose'a jest charakterystyczną cechą świadomości. Oczywiście koncepcja świadomego agenta jako wewnętrznie błędnego urządzenia, maszyny popełniającej błędy jest bezpośrednio sprzeczna ze standardowym programem (silnej) sztucznej inteligencji; co w rzeczywistości ma tendencję do wytwarzania czegoś dokładnie odwrotnego, tj. posłusznych, przewidywalnych, lojalnych maszyn, potencjalnie przydatnych do ulepszania naszych ograniczonych zdolności. Zasadniczo dzisiejsze projekty AI nadal ograniczają się do projektowania „głupich gwiazd".

19. Wewnętrzne doświadczenie

Cel eseju Mauro D'Ariano i Federico Faggina jest bardzo ambitny: autorzy przedstawiają główne kierunki teorii informacji kwantowej w pojęciu „doświadczenia wewnętrznego". Zasadniczo proponują spójną teorię, która rozwiązuje to, co D. Chalmers nazywa „trudnym problemem świadomości", a mianowicie pochodzenie i właściwości tego „wewnętrznego doświadczenia", którego każdy z nas doświadcza na co dzień. Pochodzenie qualiów (= uczuć, wrażeń) i jaźni znajduje zatem naturalne miejsce w omawianym tutaj schemacie teoretycznym. Niektóre punkty wyjścia tego eseju zostały również częściowo opisane we wcześniejszych publikacjach lub rozmowach Faggina i D'Ariano. Na przykład autorzy uważają, podobnie jak Penrose, że „prawdziwa" inteligencja wymaga świadomości, czegoś, czego nasze maszyny cyfrowe nie mają i nigdy nie będą mieć. Autorzy ci, podobnie jak Penrose, sprzeciwiają się również standardowemu postrzeganiu ludzi przez sztuczną inteligencję jako swego rodzaju „wetware". Przeciwstawiają się silnemu przekonaniu sztucznej inteligencji, że świadomość wyłania się wyłącznie z mózgu, jako produkt czegoś podobnego do oprogramowania naszych komputerów, a także poglądowi fizykalistycznemu, że świadomość „wyłania się z funkcjonowania", jak jakaś biologiczna właściwość życia. Wręcz przeciwnie, autorzy utrzymują, że podstawową właściwością

świadomości jest zdolność, zdolność odczuwania. Oczywiście zdolność odczuwania implikuje istnienie podmiotu, który odczuwa, czyli „ja". Dlatego świadomość jest nierozerwalnie związana z jaźnią, która (lub która) odczuwa „wewnętrzne doświadczenia". Centralnym punktem dyskusji jest zatem konstrukcja teorii „quali" i rozwiązanie „trudnego problemu świadomości", a mianowicie wyjaśnienie istnienia i dynamiki qualiów. Ponadto autorzy przyjmują punkt widzenia D.Chalmersa, który twierdzi, że świadomość jest właściwością podstawową, ontologicznie niezależną od jakiejkolwiek znanej (a nawet możliwej) właściwości fizycznej. Wszystkie systemy przenoszące informację mogą być świadome; świadomość jest nieredukowalnym aspektem lub właściwością natury, a nie epifenomenem. Nie jest to zjawisko wschodzące. Do pewnego stopnia świadomość jest od początku aspektem rzeczywistości. Wizja ta prowadzi autorów (podobnie jak Chalmers) do kwalifikowanego panpsychizmu. Wyraźnie wyłaniają się tu podobieństwa z koncepcją Penrose'a-Hameroffa o „atomach" protoświadomości, które powinny być obecne wszędzie tam, gdzie we Wszechświecie występuje zapadająca funkcja falowa (nie wspominając o seweryńskiej „manifestacji świata").

20. Kwantowy panpsychizm

Aby rozwiązać trudny problem świadomości, a mianowicie kwestię wyjaśnienia naszych doświadczeń zmysłowych, cielesnych, umysłowych i emocjonalnych, w tym wszelkich strumieni myśli, główną propozycją D'Ariano i Faggina jest panpsychizm, w którym świadomość jest podstawową cechą „informacja" i fizyka superweniująca na informacji. Hipoteza prezentowanej teorii głosi, że podstawową właściwością „informacji" jest jej „przeżywanie" przez wspierający ją „system". Informacje zawarte w świadomości powinny koniecznie mieć charakter kwantowy z dwóch powodów: jej wewnętrznej prywatności i jej mocy budowania myśli poprzez splątanie stanów qualiów. Autorzy nazywają swój pogląd „panpsychizmem opartym na informacji kwantowej". Za Chalmersem utrzymują, że: „Świadomość jest bytem podstawowym, którego nie da się wytłumaczyć w żaden prostszy sposób". Ale idą dalej, realizując ogólne twierdzenia Chalmersa w wyraźnym modelu technicznym. Precyzyjnie postulują: „Świadomość jest doświadczeniem systemu informacyjnego własnego stanu informacyjnego i przetwarzania". Co więcej, postulują kwantowość doświadczenia: informacyjna teoria świadomości jest teorią kwantową. I znowu podają zasadę qualiów: „Doświadczenie składa się z ustrukturyzowanych qualiów". Qualia (fenomenalne właściwości jakościowe) są elementami składowymi świadomego doświadczenia.

21 Wybór jako zmienna kwantowa

Szczególnie intrygujące jest omówienie koncepcji wolnej woli. Podczas gdy świadomość utożsamiana jest z informacją kwantową, wolna wola wytwarza skutki publiczne, które są klasycznymi przejawami informacji kwantowej. Wolę nazywa się wolną, „jeżeli jej nieprzewidywalność przez zewnętrznego obserwatora nie może być interpretowana w kategoriach braku wiedzy". Autorzy argumentują, że jeśli „wybór" jest zmienną losową, to nie może być zmienną klasyczną, gdyż zawsze można ją interpretować jako brak wiedzy. Zatem wolna wola wyboru powinna być zmienną kwantową, której przypadkowość nie może być nigdy interpretowana jako brak wiedzy. Tutaj związek z twierdzeniem o wolnej woli Conwaya i Kochana jest jasny. Ponieważ wybór jest zmienną kwantową, jego losowość powinna mieć czysto kwantowe pochodzenie. Losowość, czyli swoboda wyborów świadomości, ma tę samą naturę kwantową, co losowość cząstki kwantowej. Elektron (lub jakakolwiek inna cząstka) ma „swobodę" wyboru, którą szczelinę przekroczy: tak naprawdę nie jesteśmy w stanie przewidzieć jej indywidualnej ścieżki i niemożność ta nie wynika z braku wiedzy. To jest dokładnie to, co stwierdza Twierdzenie o wolnej woli: stopień wolnej woli elektronu jest taki sam, jak stopień wolnej woli obserwatora zewnętrznego! Ogólnie rzecz biorąc, ewolucja świadomości jest kwantowa, zatem jej rodzaj losowości jest kwantowy i nie można jej interpretować jako braku wiedzy (żadnych ukrytych zmiennych!); jako taki jest zatem bezpłatny.

22 Wolna wola i świadomość

Model D'Ariano-Faggina opisuje kilka dalszych aspektów świadomego zachowania, wykorzystując także dość wyspecjalizowane techniki obwodów kwantowych. W szczególności uważa się, że pamięć jest zasadniczo klasyczna, a przeniesienie doświadczenia kwantowego do pamięci klasycznej i odwrotnie, pamięci klasycznej do doświadczenia kwantowego, jest szczegółowo opisane z rygorystycznymi szczegółami technicznymi. W ich podejściu wolna wola i świadomość są głęboko powiązane, umożliwiając systemowi działanie w oparciu o jego doświadczenie jakościowe, przekształcając informację kwantową w klasyczną, nadając w ten sposób moc przyczynową subiektywności. Autorzy uważają, że fizyka opisuje nieograniczoną przyszłość, ponieważ świadome podmioty nie dokonały jeszcze świadomych wyborów w ramach wolnej woli. My, autorzy, jako istoty świadome, jesteśmy współtwórcami naszego świata fizycznego. Dostrzegamy tu perspektywę idealistyczną, częściowo powiązaną z opisaną przez Severina „manifestacją świata". Oczywiście idea, że nasz świat fizyczny powstaje w wyniku naszych wyborów wolnej woli, przedstawia świat jako serię aktów czysto twórczych, opartych na pojawieniu się z niczego zdarzenia elementarnego (kwantowego) i jego powrocie do niczego. Perspektywa D'Ariano i Faggina jest z tego punktu widzenia radykalnie przeciwna perspektywie Severino. Możemy zatem podsumować stwierdzeniem, że autorzy ci w pełni podzielają wizję Heraklita, w przeciwieństwie do wizji Parmenidesa i Severina.

Penrose: Pierwsza interwencja

Cóż, to dla mnie wielka przyjemność i zaszczyt być tutaj i używam tutaj tego wspaniałego kawałka nowoczesnej technologii, który...Mam nadzieję, że zadział…[urządzenie to stary projektor do folii].

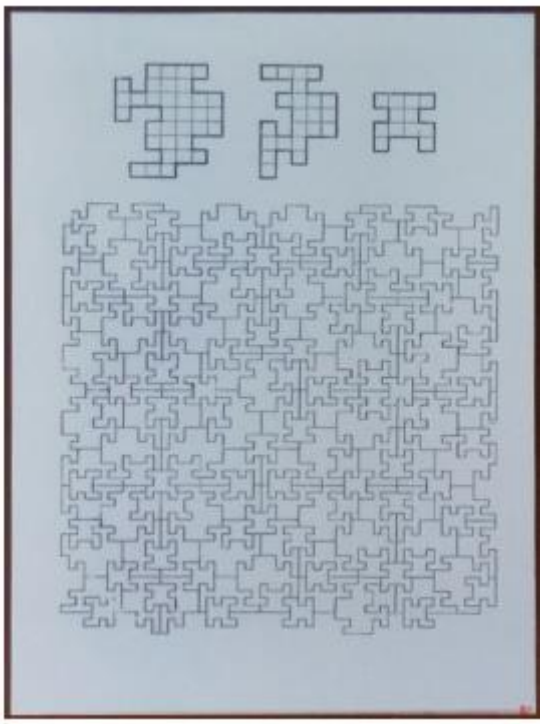
Oto tytuł, jaki nadałem tej rozmowie:

Dlaczego do zrozumienia Świadomego Umysłu potrzebna jest Nowa Fizyka.

Ta rozmowa dotyczy sztucznej inteligencji i prawdziwej inteligencji, a także tego, dlaczego moim zdaniem nie mamy teraz sztucznej inteligencji, więc przejdę do tego. Ale pozwólcie, że najpierw powiem: cóż, jest wiele rzeczy, o których nie będę mówił, jest to wiele cech świadomości, takich jak odczuwanie bólu, docenianie koloru czerwonego, miłość, intencja i wszystkie rzeczy, które są istotne. zwykle przypisywane istotom świadomym. Skoncentruję się tylko na jednej rzeczy, czyli słowie „zrozumienie". Przytoczę trzy słowa, których znaczenia nie znam. Jednym z nich jest „inteligencja", drugim „świadomość", a drugim „rozumienie". Mając wykształcenie matematyczne, nie muszę rozumieć słów, jeśli mogę powiedzieć coś o powiązaniach między nimi. Powiedziałbym więc, że w zwykłym użyciu słowo „inteligencja" wymaga „zrozumienia", i powiedziałbym, że w normalnym użyciu „zrozumienie" wymaga pewnej „świadomości". Powiedziałbym więc, że aby istota była „inteligentna" w potocznym użyciu tego słowa, musiałaby być świadoma, musiałaby być świadoma. I nie sądzę, że urządzenia, o których wiemy, są świadome. Nie zajmuję się jednak bezpośrednio kwestią świadomości. Bardziej bezpośrednio poruszę kwestię zrozumienia. Bo uważam, że zrozumienie to coś, o czym można powiedzieć coś w miarę obiektywnego. A zaczynam od pozycji szachowej, którą niedawno wymyśliłem. Wiemy, że są szachiści sterowani komputerowo, którzy grają w szachy lepiej niż jakikolwiek człowiek. Można więc powiedzieć, że z pewnością te programy szachowe rozumieją szachy. Cóż, najskuteczniejszy z tych programów szachowych nazywa się Fritz. I to stanowisko (ryc. 2) powierzono Fritzowi. Stanowisko, które wymyśliłem celowo, ponieważ jeśli zauważysz czarne figury, cóż, jest ich znacznie więcej niż białych. Możesz się martwić, że ta pozycja nie jest legalna, ponieważ jest trzech czarnych gońców, ale nie jest to nielegalna pozycja, ponieważ dwóch powstało w wyniku transformacji pionków. W tym stanowisku nie ma nic nielegalnego. To głupie stanowisko. Taka sytuacja nigdy nie miałaby miejsca w prawdziwej grze i jest to jeden z powodów, dla których Fritz nie ma pojęcia, jak grać na tej pozycji. Ponieważ dla człowieka, który wie bardzo dużo o szachach, widziałby, że dla czarnych

skoczek jest w pułapce, goniec może się trochę poruszać, król nie może się wydostać, więc wszystkie figury są uwięzione z tyłu tę barierę. Pionki również nie mogą się poruszać. A wtedy nie ma mowy o przegranej, czarne nie mogą wygrać, to oczywisty remis. Jednakże oddajesz tę pozycję Fritzowi i zakładasz, że będzie to zwycięstwo czarnych! I to, powinienem powiedzieć, jest jego upadek. Ponieważ gra przez chwilę, wykonuje całkowicie głupi ruch i przegrywa grę! Tego się spodziewałem. Przegrywa partię, bo według jakiegoś rachunku nie wiem dokładnie, co to jest, myśli, „myśli", że może to złe słowo, decyduje, że jest to zwycięstwo czarnych. Pewnie dlatego, że po prostu licząc figury, może widzi, że trzydzieści ruchów do przodu, to nie robi żadnej różnicy, wciąż ma więcej pionów i dlatego dochodzi do wniosku, że jest to zwycięstwo czarnych. Fritz jest tego zdania. Następnie zbliża się do „Zasady trzydziestu ruchów", zasada mówi: „Jeśli Twoje pionki nie zostały zabrane lub pionek się nie poruszył, oznacza to remis". A ponieważ Fritz uważa, że partia jest zwycięstwem czarnych, remis jest katastrofą! Następnie albo poświęca swojego gońca, albo zabiera gońców z linii tutaj, a król może wziąć pionka, a potem inny pionek może wkroczyć w górę i możesz zrobić trzy hetmany, całkowicie przytłaczając pozycję i jest to zwycięstwo za biały! W końcu Fritz zgadza się, że to zwycięstwo białych, ale jak widać, nie ma on żadnego pojęcia o grze w szachy, ponieważ każdy gracz, który w ogóle nie zna się na szachach, zobaczy, że to remis, i jest to bardzo głupio jest oddawać gońca lub wypuszczać pionka, który może zostać królową... Ok, to tylko szczególny przypadek. Używam tego tylko jako demonstracji, aby pokazać, że pomimo przeszkolenia, jakie mają te programy, wydaje się, że nie rozumieją gry i nie sądzę, że tak jest. Rzecz w tym, że szachy są grą skończoną, więc może zabraknąć takich pozycji jak ta. Jestem pewien, że w chwili, gdy to mówię, są ludzie, którzy próbują zaprogramować Fritza, aby – gdyby dostał takie stanowisko – nie wykonał tego głupiego ruchu. To nie jest takie proste. Ale oczywiście, że możesz! Fritz nie decyduje jednak, jak zmienić program. To człowiek, który rozumie, co się dzieje, który rozumie program leżący u podstaw zachowania Fritza. Zatem zrozumienie leży po stronie istoty ludzkiej, a nie Fritza. Niemniej jednak szachy to skończona gra, więc możesz się martwić, że ok, można je całkowicie przeanalizować i wtedy takie błędy nie będą już popełniane. Musimy więc skoncentrować się na nieskończoności. Niektórzy uważają, że nieskończoność to coś, czego nie można zrozumieć, czego nie można kontemplować. Niemniej jednak nie jest to prawdą. Możemy zrozumieć nieskończoność. Każdy z Was jest w tym dobry. Jeśli otrzymasz dwie liczby parzyste i dodasz je do siebie, otrzymasz kolejną liczbę parzystą. To jest stwierdzenie dotyczące wszystkich liczb. Możemy to zrobić. Ale to jest stwierdzenie dotyczące nieskończonej liczby rzeczy. Nie ma problemu z mówieniem o nieskończonej liczbie rzeczy i zrozumieniem nieskończonej liczby rzeczy. Czasami może to być bardzo trudne. Czasami może to być łatwe. Twierdzę, że maszyny nie mają takiej możliwości. Istnieją twierdzenia, które mniej więcej to mówią. Podam znane twierdzenie. Podam ci twierdzenie Goedela. Nie, poczekaj chwilę, najpierw chcę ci opowiedzieć, w jaki sposób uczymy się o nieskończoności, kiedy idziemy do szkoły. Uczymy się indukcji matematycznej. Tylko w ten sposób możemy zrozumieć nieskończoność. Mówię o liczbach naturalnych, które są wszystkimi nieujemnymi liczbami całkowitymi. Załóżmy teraz, że masz propozycję P(n), która zależy od liczb naturalnych. Teraz, jeśli udowodnisz, że to twierdzenie jest prawdziwe dla zera lub jeśli możesz zobaczyć, że jest ono prawdziwe dla zera i możesz udowodnić, że jeśli jest prawdziwe dla n, to jest prawdziwe dla n + 1, to wiesz, że jest prawdziwe dla wszystkich liczb ! Samo ustalenie dwóch rzeczy, tych dwóch rzeczy, mówi ci, że twierdzenie jest prawdziwe dla wszystkich liczb. W ten sposób można ustalić, co jest prawdziwe dla wszystkich liczb. Ale powiem coś o tym ograniczeniu. Po pierwsze, stwierdzenie bardzo ogólne, które wynika z wypowiedzi Goedla. Ale podam wam to w formie Turinga, co jest bardzo ładnym sposobem wyrażenia tego. Załóżmy, że masz zbiór reguł, nazywam go R, czyli dowolny zbiór reguł, który można sprawdzić. Widzisz, masz zestaw twierdzeń, z których każde wynika z poprzedniego na podstawie reguł R. Ja tylko upieram się, że komputer mógł sprawdzić, czy reguły są poprawnie stosowane, czy nie. Załóżmy teraz, że te reguły mają służyć dowodzeniu twierdzeń o liczbach. Może to być ostatnie twierdzenie Fermata udowodnione przez Andrew Wilesa lub hipoteza Goldbacha, która mówi, że każda liczba parzysta

większa niż 2 jest sumą dwóch liczb pierwszych, która wciąż nie została udowodniona. Właśnie o takie propozycje mi chodzi. Twierdzenia dotyczące nieskończonej liczby rzeczy. Zasady muszą być takie, że jeśli zastosujesz się do nich poprawnie, czyli podasz im określone stwierdzenie, a zasady odrzucą i jeśli uda im się uzyskać „Tak, to prawda", ok, to musisz uwierzyć, że tak jest PRAWDA. Mogą powiedzieć „nie, to fałszywe", ok, albo mogą nic nie powiedzieć. Mogą po prostu trwać i trwać, i trwać, i trwać w nieskończoność. Załóżmy teraz, że wierzysz, że zasady nigdy nie zawierają fałszywego stwierdzenia, na przykład 2 plus 2 równa się 3. Załóżmy, że nigdy nie dają ci czegoś, co jest rażąco fałszywe, a jedynie dają ci rzeczy, które są prawdziwe. Następnie pokażemy, że można skonstruować określone stwierdzenie dotyczące liczb, takie jak hipoteza Goldbacha, jak twierdzenie Fermata, bardzo konkretne stwierdzenie, a ze sposobu, w jaki je skonstruowałeś, możesz zobaczyć, że: po pierwsze, stwierdzenie jest prawdziwe; po drugie, oświadczenia nie można uzyskać za pomocą przepisów. To jest bardzo niezwykła rzecz. Kiedy usłyszałem to po raz pierwszy, byłem oszołomiony. Ponieważ nie powiedział ci, że tych rzeczy nie można udowodnić, ponieważ sama rzecz, którą konstruujesz, jest czymś, o czym wiesz, że jest prawdziwa. Ale wiesz, że to prawda tylko dlatego, że ufasz zasadom. Oznacza to, że jeśli wierzysz, że zasady podają jedynie prawdy, to twoja wiara wykracza poza same zasady. Bardzo niezwykła rzecz! Oznacza to, że zrozumienie, dlaczego reguły działają, mówi ci coś, czego reguły nie mogą osiągnąć. Mówi ci, że zasady nie mogą czegoś osiągnąć, ale mimo to coś jest prawdą. Możesz zobaczyć, że to prawda, po sposobie, w jaki to skonstruowałeś, jeśli ufasz zasadom. Kiedy po raz pierwszy o tym usłyszałem, byłem zdumiony! Ponieważ wydaje się nam mówić, że jest coś wyjątkowego w naszym rozumieniu – nie musisz tego wymyślać, nie musisz być Goedelem, nie musisz być Turingiem, nie musisz mieć inspiracją, aby zobaczyć, dlaczego to musi być prawda. Wszystko, co musisz zrobić, to postępować zgodnie z rozumowaniem, a podążając za rozumowaniem, zobaczysz, że stwierdzenie wynikające z G(R) jest prawdziwe, jeśli ufasz regułom. Zrozumienie, dlaczego zasady działają, daje ci coś, czego zasady nie dają. Cóż, ten fakt wyraźnie mówi, że zrozumienia nie można osiągnąć za pomocą zasad. Teraz jest na ten temat coś więcej, o czym za chwilę. Ale pozwólcie, że podam wam kilka przykładów rzeczy, których komputery nie potrafią. Takich przykładów jest wiele. Jest to miłe, ponieważ jest bardzo łatwe do zrozumienia. Załóżmy, że otrzymałeś coś takiego (ryc. 3), które jest zrobione z kwadratów tej samej wielkości i sklejonych ze sobą wzdłuż krawędzi, dzięki czemu możesz tworzyć różne kształty na płaszczyźnie. Nazywa się je poliomino. Może się zdarzyć, że niektórymi z tych kształtów uda się pokryć całą płaszczyznę euklidesową, po prostu używając tego kształtu. Oto przykład, dzięki któremu można pokryć całą płaszczyznę. A tutaj jest jeszcze jeden, który pozostawia pewne luki pomiędzy kształtami i tak dalej. Istnieje twierdzenie matematyczne autorstwa Roberta Bergera, które stwierdza, że nie ma algorytmu komputerowego, który powie „tak" lub „nie", czy skończony zbiór poliomino ułoży płaszczyznę obok siebie. To bardzo niezwykłe. Pozwól mi dać ci przykład. Oto przykład



Sam go wymyśliłem, ale pokazuje on trudności. Te trzy kształty pokryją płaszczyznę. Ale jeśli spojrzysz na ten wzór i będziesz się mu przyglądać przez chwilę, czy zobaczysz wzór, według którego jest zbudowany? Bardzo znany matematyk powiedział mi: „Nie rozumiem, jak to jest zbudowane. Jak byś to zrobił? Czy istnieje algorytm?". Cóż, istnieje algorytm, który pozwala zobaczyć, jak to zrobić. Musisz zajrzeć do moich dzieł zebranych. Jest jeden dokument, w którym można znaleźć, jak to zrobić. Ale nie wolno ci przeglądać moich papierów, to oszustwo! To nie jest rodzaj algorytmu ani procedury, mam na myśli! Niemniej jednak istnieje sposób na pokrycie tymi kształtami całej płaszczyzny, ale nie znam żadnego programu komputerowego, który byłby w stanie to zrobić! Podałem to tylko jako przykład czegoś, czego nie da się obliczyć. Można to jednak zrozumieć, stosując odpowiedni sposób rozumowania. Może nie wiem, czy tak jest w każdym przypadku. Wydaje się jednak możliwe, że na podstawie wglądu można stwierdzić, przynajmniej w wielu przypadkach, czy te rzeczy można zrobić. Ale nie ma na to ogólnego algorytmu. Podam inny przykład z tym związany. To jest przykład, który podał w 1945 roku matematyk Reuben Goodstein, a ja chcę spróbować powiedzieć, co to jest (ryc. 5). Bierzesz dowolną liczbę naturalną, tutaj biorę jako przykład 1077. Najpierw zapisujesz tę liczbę w formacie binarnym. Reprezentujesz tę liczbę jako sumę różnych części zapisu binarnego. Zapisujesz także wykładniki binarnie. Teraz poddam ci dwie operacje. Jedna operacja to A. A mówi: spójrz na wszystkie „dwójki" i zamień je na „trójki". OK, dobrze zdefiniowana operacja. Liczba ta staje się znacznie większa, nieważne. To jest operacja A. Następnie operacja B: odejmij 1. OK, to tylko trochę się sprowadza. Z drugiej strony operacja A: zamień wszystkie „trójki" na „czwórki". Następnie Operacja B: odejmij 1. To niesamowity przykład zająca i żółwia. Żółw zawsze wygrywa. To niezwykłe. Co się dzieje z tymi liczbami? Pierwsza to 1077, następna około 1040, następna 10600... i coraz większa, i większa... aż w końcu wychodzi zero! Bardzo niezwykłe! Ponieważ liczba staje się coraz większa, ale odjęcie jednego (żółwia) na końcu zawsze wygrywa. Jest to niezwykłe nie tylko dlatego, że jest prawdziwe, ale dlatego, że J. Paris i L. Kirkby (w 1982 r.) udowodnili, że nie można udowodnić tego twierdzenia za pomocą zwykłej indukcji matematycznej. To jest przykład twierdzenia Goedela. Aby to udowodnić, należy zastosować tzw. indukcję pozaskończoną. Sugeruję, jeśli chcesz spróbować, spróbuj z 3 lub 4. Nie polecam umieszczania go na swoim laptopie. Nie polecam korzystania z komputera typu mainframe. Nie polecam korzystania z żadnego komputera. Sugerowałbym, abyś użył ołówka i papieru

i prawdopodobnie możesz przekonać samego siebie, że ostatecznie spadnie do zera. Ale jeśli umieścisz to na jakimkolwiek komputerze, który istnieje dzisiaj na świecie lub jest możliwy do pomyślenia na świecie, nigdy tego nie zrobi. Liczba ta staje się tak ogromna, że jest to niezwykłe... nawet przy czterech... Ale możemy to zrozumieć. Nie jest to takie oczywiste, ale możemy. Skąd to wiemy? Jak to jest możliwe? Niektórzy mówią, że mamy w głowie algorytm, który jest tak skomplikowany, że nie możemy zastosować twierdzenia Goedla. I dlatego to nie działa. Nie wierzę w to. Jednym z powodów, dla których w to nie wierzę, jest to, że... jak doszło do naszego zrozumienia? Cóż, przez dobór naturalny! W to wierzę! Ale dobór naturalny nie jest zbyt dobry w tworzeniu matematyków. Zrozumienie w sensie ogólnym, tak, rzeczywiście, było bardzo ważne. Ale zrozumienie matematyki? Nie bardzo! Oto matematyk . Prawdopodobnie ma w głowie algorytm, ale zostanie zjedzony przez tygrysa szablozębnego! ...Chcę powiedzieć, że bycie matematykiem nie jest tak naprawdę selektywną zaletą! Jak więc to się dzieje, że mamy zrozumienie? Dlaczego możemy robić rzeczy, których nie potrafi żaden wyobrażalny komputer? (Widzieliśmy to w przypadku twierdzenia Goodsteina.) Jaka jest ogólna jakość zrozumienia? Co nie jest algorytmem! Zrozumienie nie jest elementem bardzo skomplikowanego algorytmu. To wszystko, co próbuję powiedzieć. Jaka jest zatem selektywna zaleta ogólnej zdolności rozumienia? I nie jest to jakość obliczeniowa. To jest przesłanie, które próbuję przekazać. Kiedy byłem studentem w Cambridge, poszedłem na kurs logiki matematycznej. Tam dowiedziałem się o Goedlu i Turingu i byłem zdumiony. Ale potem pomyślałam: ok, skoro naszymi głowami nie rządzą algorytmy, żeby wyglądały jak komputer, to co to może być? Próbowałem zająć się fizyką. Próbowałem zrozumieć, jak działa świat fizyczny. I zrozumiałem, że można umieścić w komputerze bardzo skomplikowane rzeczy, takie jak równania ogólnej teorii względności Einsteina, i symulować takie rzeczy, jak łączenie się czarnych dziur i emisja fal grawitacyjnych. Ludzie wykonali bardzo skomplikowane obliczenia, aby określić dokładny kształt sygnału fal grawitacyjnych, który wskazuje, że para czarnych dziur zbliża się do siebie. Obliczenia! Za pomocą obliczeń możesz symulować działanie teorii Einsteina z ogromną precyzją. Ok, więc co możemy zrobić w naszych głowach? Może to nie jest teoria względności, może to zwykła fizyka Newtona. Cóż, dzięki temu można wykonać bardzo, bardzo dobre obliczenia. A co z mechaniką kwantową? Cóż, to interesujące pytanie. Ponieważ możesz postępować zgodnie z równaniem Schroedingera; znowu może to być dość trudne, ale znowu ma charakter czysto obliczeniowy. Kiedy byłem studentem, poszedłem na kurs prowadzony przez wielkiego fizyka Paula Diraca i już na pierwszym wykładzie miał kawałek kredy, który położył w dwóch miejscach. Mówił o zasadzie superpozycji w mechanice kwantowej, która mówi, że rzecz, która może znajdować się w superpozycji stanów, może znajdować się w dwóch miejscach na raz, z pewnymi liczbami zespolonymi jako funkcjami ważącymi. A potem Dirac wyjaśnił, dlaczego kawałek kredy nie może być w dwóch miejscach na raz, a mój umysł niestety zabłądził i nigdy nie usłyszałem wyjaśnienia! :) Od tamtej pory się tym martwię! Powiedział coś niejasno o energii, ale nie wiem, co to było. Od tego momentu martwiłem się, jak na Ziemi kawałek kredy nie może znajdować się w superpozycji. Bo zgodnie z zasadami mechaniki kwantowej w zasadzie można w superpozycji umieścić kawałek kredy. Schroedinger pokazał, że według mechaniki kwantowej kot może być jednocześnie martwy i żywy. Powinienem podkreślić, że opracował ten przykład tylko po to, aby pokazać, że jego własne równanie nie może być całą historią. Musi być coś innego. Celowo pokazywał absurdalność własnego równania. Myślę, że to bardzo niezwykłe! Teraz opowiem wam o mechanice kwantowej. Używam tego rysunku do zilustrowania mechaniki kwantowej. Ten obraz powstał dlatego, że zostałem zaproszony do wygłoszenia wykładu przez Chrześcijańskie Towarzystwo Andersena w Kopenhadze, a stało się to dzięki napisaniu książki pt. Nowy umysł cesarza, będącej sztuką na temat Nowe szaty cesarza Andersena. To zdjęcie jest częścią wykładu, który wygłosiłem i ilustruje mechanikę kwantową. Jak może zilustrować mechanikę kwantową? Obraz składa się tak naprawdę z dwóch rzeczy. Góra to klasyczny świat. Widzisz oddzielne rzeczy, budynki, ludzi wokół, mają lokalizację, tu i tam, itd. Dno to szalenie wyglądający świat kwantowy, który jest splątany w skomplikowane sposoby i to wszystko znajduje się pod powierzchnią

morza. A jaka jest rola syreny? Cóż, syrena jest magiczna. Pokazuje, jak przejść z jednego świata do drugiego. Jest magią redukcji stanu, czyli załamania funkcji falowej. To właśnie trzeba robić w mechanice kwantowej. Podążasz za równaniem Schroedingera do pewnego momentu, a potem mówisz: cóż, nie, to robi się zbyt szalone... martwe koty i żywe koty. Ach, spójrz do pudełka, a twoja świadomość jest w to zaangażowana i w jakiś sposób sprawiasz, że kot jest martwy lub żywy, zamiast znajdować się w superpozycji. Nigdy w to nie wierzyłem. Wydaje się, że istnieje coś w fizyce świata, co dokonuje takiego wyboru, a jest to magia załamania się funkcji falowej lub redukcji stanu kwantowego, co właśnie robi syrena. W ten sposób fizyka klasyczna wychodzi ze świata kwantowego. Nie objawia się to w sposób, w jaki obecnie rozumiemy mechanikę kwantową. Stopniowo zacząłem więc myśleć, że to, czego nie zrozumiałem podczas wykładu Diraca, jest naprawdę głębokie. Czegoś brakuje w fizyce. I to jest jedyna rzecz, jaką widziałem, której nie można umieścić na komputerze! Dlatego wydawało mi się, że właśnie z tego możemy skorzystać. Z czego składa się świadomość. Pozwólcie, że zilustruję teraz tę tezę przykładem, wzorem, który opracowałem po długim namyśle (węgierski fizyk L. Diosi wymyślił podobny wzór wcześniej ode mnie) . Załóżmy, że masz ciało i próbujesz ułożyć je w dwóch pozycjach jednocześnie i chcesz wiedzieć, jak długo to potrwa. Teraz mówimy, że stanie się jednym lub drugim w określonym czasie, który zależy od energii. Nie wiem, co dokładnie Dirac miał na myśli, ale wystarczy spojrzeć na energię w tej konfiguracji i energię w tamtej, a także na rozkład masy i odjąć jedno od drugiego. Następnie oblicza się tak zwaną energię własną grawitacji tego układu o dwóch konfiguracjach, mianowicie Eg, i na koniec czas życia tej superpozycji stanów jest określony wzorem ħ/Eg. Istnieje więc wzór, który powie ci, kiedy utwór Diraca kredy trafiłoby tutaj, a kiedy znalazłoby się tam. Formuła nie powie Ci gdzie, ale powie Ci, kiedy stanie się jedną lub drugą konfiguracją. Pomysł jest taki, że tego procesu nie da się obliczyć. W rzeczywistości jest to coś, co przypomina trochę wolną wolę. Ponieważ zgodnie z obecną fizyką podejmuje własne decyzje. W jakiś sposób decyduje się być tu czy tam, jakoś to decyduje... Ale nie wiem, co to znaczy! Tak więc wraz z moim kolegą Stuartem Hameroffem opracowaliśmy teorię, w której staramy się powiedzieć, że świadome decyzje są właśnie takimi procesami. „Załamanie" w jednym lub drugim stanie. Miałem takie pomysły, kiedy pisałem Nowy umysł cesarza, ale nie miałem pojęcia, gdzie w mózgu mogło wydarzyć się coś takiego. Zaczynam czytać i myśleć o przekazie nerwowym, o tym i tamtym, i tak dalej, ale to nie może być, to nie może być... I pomyślałam optymistycznie: „pod koniec książki poznam odpowiedź". To było głupie z mojej strony. Pod koniec książki nie znałem odpowiedzi. Ale mimo to nie przestałem pisać tej książki, została ona opublikowana i miałem nadzieję, że książka może zainspirować niektórych młodych ludzi. Ale pisali do mnie głównie starsi ludzie, którzy przeszli na emeryturę. Być może tylko oni mają wystarczająco dużo czasu, aby przeczytać tę książkę. W każdym razie dostałem list od Stuarta Hameroffa, którego zadaniem jest uśpienie ludzi w odwracalny sposób. Jest anestezjologiem. Jednak w przeciwieństwie do wielu swoich kolegów chce wiedzieć, co naprawdę robi. Oczywiście chce uśpić ludzi, aby nie narazili ich na niebezpieczeństwo i aby można było ich ponownie obudzić, nie narażając się na konsekwencje wykonanej operacji. Operacja zakończyła się sukcesem, ponieważ osoba była nieprzytomna. Stuarta bardzo interesowało to, co właściwie robi. Wpadł na pomysł, że znieczulenie ogólne nie wpływa bezpośrednio na neurony, ale na mikrotubule . A kiedy o tym usłyszałem, pomyślałem, że to dużo bardziej prawdopodobne. Ponieważ są to długie, cienkie nanoskopowe rurki, które wydają się zachowywać spójność kwantową, a tego właśnie chcemy. Wypracowaliśmy między sobą teorię. Nadal nie wiemy, co się naprawdę dzieje, ale mikrotubule są tego kluczowym elementem, a twierdzi się, że wewnątrz tych małych rurek dzieje się coś z mechaniki kwantowej. Wpływają na siłę synaps i dlatego umieszczają je na głębszym poziomie niż synapsy neuronowe. To, co sprawia, że struktura neuronu robi to, czego chcesz, jest kontrolowane na głębszym poziomie, a kontrolowanie tego ma związek z nową fizyką mikrotubul. Tego jeszcze nie wiemy, ale sądzę, że jest to coś w rodzaju procesu, o którym przed chwilą mówiłem. Jest więc wiele ekscytujących rzeczy do odkrycia tutaj. Czy to prawda? Uważam, że istnieją pewne dobre dowody, najnowsze dowody na to, że działanie środków do

znieczulenia ogólnego polega po prostu na tym, że działają na mikrotubule. Mikrotubule wydają się mieć kluczowy wpływ na działanie środków do znieczulenia ogólnego. Myślę, że to obiecujące. Ale to, co się dzieje, jest przyszłością i myślę, że jest naprawdę interesujące. Jednakże ten punkt widzenia sugeruje, że w świadomym działaniu kryje się coś zupełnie innego. Cóż, w terminologii, którą wymyśliliśmy, mówimy „protoświadomość". Za każdym razem więc, gdy świat podejmuje decyzję, a mianowicie za każdym razem, gdy następuje załamanie się funkcji falowej (syrena!), wówczas wiąże się to z elementem protoświadomości. To jeszcze nie jest prawdziwa świadomość, bo nie ma celu, nie ma mechanizmu, który nadaje temu sens. Pomysł jest jednak taki, że są to elementy składowe, z których zbudowana jest świadomość. Kiedy to zrozumiemy, być może stworzymy świadome urządzenie. Jednak dzisiejsze komputery nie są świadome i dopóki pozostaną komputerami, pozostaną nieświadome. To mówi nam, że nie przejmą władzy nad światem. Nie sądzę, żeby to było niebezpieczeństwo. Myślę, że istnieją niebezpieczeństwa, jest wiele niebezpieczeństw, nie mówiłem o tym i nie mam teraz czasu, aby o tym rozmawiać. Więc nie. Pozostawiam to innym do dyskusji na ten temat. Dziękuję bardzo.

Severino: Pierwsza interwencja

Szczególnie gratuluję błyskotliwej i głębokiej komunikacji profesora Penrose'a. Z pewnością nie jest rzeczą właściwą, aby filozofia wkraczała w naukę. Nauka bardzo dobrze wie, co zamierza zrobić. Ale filozofia może próbować zrozumieć znaczenie słów, którymi posługuje się wiedza naukowa, a profesor Penrose użył ważnych w filozofii słów, w szczególności słów „świadomość", „inteligencja" i „rozumienie". Co więcej, powiedział niezwykle interesujące rzeczy na temat różnicy między tym, co może zrobić komputer, a tym, co może zrobić świadoma istota. Jednak myśl filozoficzna, która dziś niemal pozostaje w cieniu wielkości wiedzy naukowej, popycha myśl ludzką tak daleko, jak to tylko możliwe. Przede wszystkim zaczyna się od stwierdzenia, że nauka nie może głosić prawd niezaprzeczalnych. Ale nie tylko filozofia to mówi; to sama nauka rozpoznaje siebie jako zbiór wiedzy hipotetyczno-dedukcyjnej, probabilistycznej i uznają to nawet nauki matematyczne. Filozofia z pewnością nie powinna więc mieszać się w nauce, bo jak teraz wspomnę (ale będzie to tylko podpowiedź), nauka podsumowując nie dąży do prawdy, ale do władzy nad światem. A o wyborze między dwiema różnymi i konkurującymi teoriami ostatecznie decyduje zdolność jednej, a nie drugiej, do przekształcenia świata. To sama nauka uznaje, że nie jest już możliwe robienie tego, w co Galileusz wciąż wierzył, że jest w stanie dokonać, czyli budowanie wiedzy, której nikt, ani bogowie, ani ludzie, nie może zaprzeczyć. Nauka dąży do zdobycia władzy nad światem. I nawet matematyka, nawet jeśli tylko niewielka część możliwości matematyki jest faktycznie wykorzystywana w polu fizycznym, o ile wiem, nawet matematyka ostatecznie nie dąży do prawdy absolutnej. Weźmy na przykład Principia Mathematica, z którego korzystał Goedel budując swoje twierdzenia, w którym mieści się twierdzenie wspomniane przez prof. Penrose'a. Principia Mathematica, nad którymi zastanawia się Goedel, również wychodzi od postulatów. Postulowanie oznacza, że proszę Cię o przyznanie mi czegoś, z czego następnie, zgodnie z pewnymi regułami, musi wynikać coś innego. Nie sądzę jednak, aby nawet te zasady mogły pochwalić się zasadą niepodważalną; to znaczy, że same są konwencjami. Zatem zgadza się, że po danym twierdzeniu musi nastąpić pewne inne twierdzenie, dokładnie według pewnej reguły. Ale to, że reguła jest całkowicie bezsporna, nie jest nawet tym czymś bezspornym. Druga obserwacja dotycząca tego, co powiedział profesor Penrose (ale którą mógłbym również rozszerzyć, jak to zrobiłem wcześniej w odniesieniu do nauki, na całą dziedzinę wiedzy naukowej) jest taka, że gdy nauka mówi o świadomości, inteligencji lub rozumieniu, nauka odnosi się do konkretnych obiektów , do konkretnych rzeczy. Zrozumienie tego, zrozumienie tamtego, dokonanie tej operacji, dokonanie tej innej operacji.Ale kiedy mówimy o „szczególnych wymiarach", czyli świadomości rozumianej jako „szczególna rzecz", cóż, skąd dokładnie to mamy wiadomość o „świadomości"? Z czego dokładnie czerpią nauki neurologiczne, fizyczne i psychologiczne? Z jakiego źródła czerpie się informacje o

istnieniu świadomości, świadomym działaniu lub fakcie psychicznym? Czerpią go z wymiaru, na który nauka na ogół nie zwraca uwagi, a jeśli chcemy powiedzieć, czym jest ten wymiar, musimy powiedzieć, że jest to „przejaw świata". Nauka nie może podjąć żadnego kroku, który nie zaczynałby się od manifestacji świata. Ale to nie jest rzeczą pośród rzeczy, ponieważ przejaw świata obejmuje także rzeczy przeszłe i przyszłe. Kiedy na przykład mówimy o teorii ewolucji, ewolucja świadomości jest czymś, co leży w szerszej świadomości, z której wyłania się coś, na przykład ewolucja umysłu, ewolucja człowieka... Zatem, kiedy mówimy o wytwarzaniu świadomości, mamy na myśli należy zwrócić uwagę, aby podkreślić, że mamy na myśli produkcję lub wytwarzalność konkretnego przedmiotu, ponieważ przejaw świata jako całości nie może być przedmiotem produkcji. Przynajmniej z tego powodu: ponieważ producent, gdyby miał wyprodukować całość przejawu świata, znajdowałby się poza przejawem świata, a zatem byłby czymś nieznanym. Trzecia uwaga dotycząca zaproponowanego przez organizatorów tematu, dotyczącego relacji pomiędzy sztuczną inteligencją a inteligencją naturalną. Chcemy wytwarzać inteligencję, ale kiedy zastanowimy się, czym jest inteligencja, musimy także zająć się tym, co myśl filozoficzna powiedziała na temat inteligencji. Prawdą jest, że w fizyce i w ogóle w przypadku wszystkich nauk istnienie świata niezależnego od inteligencji jest czymś powszechnie akceptowanym. Nie wolno nam jednak zapominać, że ten świat, w założeniu istniejący poza inteligencją, jest znowu czymś, co rozumiemy, a zatem nie jest czymś, co leży poza tym „rozumieniem", które poprzednio nazwałem „przejawem świata". I nie należy zapominać, że w odniesieniu do tej koncepcji świata, który istnieje niezależnie od świadomości, myśl nowoczesna wskazała właśnie, że ten świat nie jest światem istniejącym poza świadomością, ponieważ ten świat jest światem poznanym i ponieważ jak wiadomo, jest treścią świadomości i dlatego nie może znajdować się poza świadomością. Aby zrozumieć tę kwestię, wystarczy zdawać sobie sprawę z tego, co powiedział na ten temat Kartezjusz: Kartezjusz zdefiniował „ten świat", cytuję z pamięci, „ea omnia quatenus in nobis sunt et in nobis fiunt et in nobis eorum conscientia est" czyli „wszystko, co w nas jest, staje się w nas i jesteśmy tego świadomi". Z biegiem czasu dualistyczna koncepcja inteligencji i świadomości została wyparta. Ale nie będziemy o tym dyskutować, bo to by nas za bardzo oddaliło, a zamiast tego skupimy uwagę na cesze decydującej nie tylko w odniesieniu do historii nauki, nie tylko w odniesieniu do historii kultury, ale ja powiedziałbym w odniesieniu do historii ludzkości. Wydaje mi się, że nawet profesor Penrose użył wyrażenia „budowanie świadomości" lub „wytwarzanie świadomości"... a w każdym razie chodzi tu o możliwość wytwarzania świadomości. Tutaj terminem, który jest oczywisty i nigdy nie kwestionowany, ale który w rzeczywistości jest decydujący, jest właśnie termin „produkcja". Dobrze wiemy, że wszystkie nauki, nie tylko przyrodnicze nauk ścisłych, ale także nauk historycznych, używają tego terminu, nie zadając sobie trudu dostrzeżenia otchłani, jaka kryje się pod tym terminem. Produkować. I nawet tutaj, zanim wspomniałem o Kartezjuszu, ale pamiętam Platona i przypominam sobie definicję produkcji, którą podaje Platon, i produkcji, ponieważ jest to definicja, która ostatecznie pozostaje u podstaw wszystkiego, czego dokonał Zachód, pomyślał: lub realizowane w obszarze naukowym i pozanaukowym. Jeżeli teraz wspomnę tę platońską definicję produkcji, wielu z Was zapewne powie: cóż, my to wszystko wiemy! Oczywiście wiemy o tych rzeczach, ponieważ stały się tym, o czym bez wątpienia jest obecnie przekonana nasza cywilizacja oczywiste, więc stało się to zdrowym rozsądkiem i jeśli ktoś tak jak ja zaprasza nas do refleksji nad tym, o czym wszyscy wiemy, to dziwimy się i naturalnie mówimy: nie powinniśmy tracić czasu na te oczywiste rzeczy! Przejdźmy raczej do szczegółów, na przykład wiedzy naukowej! A co Platon mówi o produkcji? Nazywa to „poiesis". Tutaj także cytuję z grubsza, jeśli pozwolicie mi najpierw powiedzieć tekst grecki, a następnie go przetłumaczyć, ale tekst grecki zacytuję w małych fragmentach. Produkcja to „poiesis". Mam na myśli fragment Convivium, który jest fragmentem 205 p.n.e. (Convivium lub Sympozjum). Platon powiada: wytwarzanie to dowolna przyczyna – nazywa ją „aitia" – jest to każda przyczyna, która sprawia, że coś przechodzi z niebytu do bycia „hec tu me on, eish do on". Produkcja jest przyczyną, która sprawia, że wszystko przechodzi z niebytu do bytu. Od tego czasu nie było niczego w myśli zachodniej, ani na polu

naukowym, ani pozanaukowym, gdzie ta obserwacja, obecnie uważana za banał, nie posłużyła za podstawę. Jest podstawą wszystkiego, co myślimy i robimy. Stanowi także podstawę pojęcia produkcji. Więc nawet jeśli chcemy porozmawiać o produkcji świadomości, nie możemy uniknąć rozważenia znaczenia produkcji, „poiesis". Produkcja jest przyczyną, która prowadzi od niebytu do bytu. W tym fragmencie Convivium Platon mówi dalej, że wszystkie „ergasiai", czyli czynności wykonywane w technikach, czyli „technai", są wytworami. Tak samo „demiurgoi", ci, którzy produkują na różnych polach, są producentami; nazywa je „poietai". W tym miejscu pojęcie „technologii" zostaje wprowadzone w powiązaniu z pojęciem „produkcji". Teraz wszyscy jesteśmy przekonani, że nie ma sensu mówić o tak elementarnych pojęciach, które uważamy za absolutnie oczywiste, i wszyscy jesteśmy o tym przekonani, zwłaszcza po wysłuchaniu tak interesujących szczegółów, jak te wskazane przez profesora Penrose'a. Chciałbym jednak zwrócić uwagę i tu dochodzimy do sedna mojego wywodu, że to, w co wszyscy wierzymy i co nauka (każda nauka) również uważa za niezwykle oczywiste, a mianowicie poiesis, przejście od niebytu do bytu , to coś, za co wszyscy uważamy niezwykle oczywiste, w rzeczywistości nie jest obserwowalne. I to na początku z pewnością rozczaruje znaczną liczbę dzisiejszych słuchaczy, jeśli nie większość z nich. Bo na pewno będą protestować w związku z tym twierdzeniem: Co masz na myśli? Przejście od niebytu do bycia jest czymś oczywistym. Rodzimy się, umieramy, poruszamy przedmiotami. Teraz mówisz nam, że tego nie da się zaobserwować! Ma to istotne implikacje, gdyż wówczas każdy rodzaj produkcji, a zatem także możliwa produkcja świadomości, jest czymś, czego nie da się zaobserwować. Pozwolę sobie bardzo szybko wskazać przyczynę twierdzenia o przejściu od niebytu do bytu lub od bytu do niebytu. Podam przykład: miasto Hiroszima zostało zniszczone. Kiedy Hiroszima została zniszczona, nie wyglądała już tak, jak wyglądała przed zniszczeniem. Tego już nie dało się zaobserwować. Hiroszimę można pamiętać, ale nie można jej już obserwować taką, jaką była, gdy nie została jeszcze zniszczona. A zanim został zniszczony, był zauważalny. Dlatego zniszczenie dla tych, którzy wierzą, że to zniszczenie było przejściem miasta od bytu do niebytu, musi być jednocześnie formą nieprzynależności już do przejawu świata. Chodzi o to, aby pomyśleć o tej idei: jeśli wierzysz, że coś było niczym, to zaczyna być, a potem znowu staje się niczym, cóż, to stwierdzenie jest teorią, a nie obserwacją. Dla przykładu moglibyśmy powiedzieć, że widzimy Słońce na sklepieniu nieba, a po zachodzie słońca już go nie widzimy, tak jak nie widzieliśmy go przed wschodem o świcie. Ale czy ma sens wierzyć, że sklepienie nieba, czyli to, co można zaobserwować, ukazuje nam losy Słońca przed wschodem i po zachodzie? NIE! Niebiańskie sklepienie nie ukazuje losów Słońca, kiedy nie znajduje się ono jeszcze w skarbcu i kiedy go już nie ma. Chciałbym poczynić jeszcze jedną obserwację. Odniesienie profesora Penrose'a do Gödela było bardzo interesujące. Istnieją prawdy, mówi Goedel, których nie da się udowodnić, ale są prawdziwe. I to są prawdy, które goedeliści (ale zdaje mi się, że sam Gödel to mówił) czasami nazywają prawdami intuicyjnymi. Ale także w tym przypadku, co oznacza „prawda"? Teraz jest jasne, że logika i matematyka używają słów „prawda" i „nieprawda". Używają jednak słów, których znaczeniu powinniśmy się bardzo uważnie przyjrzeć. Jedną z prawd jest na przykład manifestacja świata o której wspomniałem wcześniej. I jesteśmy przekonani, że jesteśmy tutaj, w Cariplo, w Mediolanie, we Włoszech, z prof. Penrose'em. Ale temu przekonaniu, które jest intuicją, przejawem świata, tą intuicją, dlaczego nie można temu zaprzeczyć? No cóż, ktokolwiek temu zaprzecza (a osobiście jestem przekonany, że jestem tutaj, w Mediolanie, w Cariplo), ten, kto temu zaprzecza, uważany jest za wariata i umieszczany jest w szpitalu psychiatrycznym (choć dziś nie ma już szpitali psychiatrycznych, więc powinniśmy powiedzieć: zostałby umieszczony w szpitalu psychiatrycznym). Umieszczenie kogoś w ośrodku psychiatrycznym nie oznacza jednak, że znajduje się on w nieprawdzie. Oznacza to, że wyraża się ktoś, kto jest określany mianem szaleńca. Innymi słowy, to, co dla Goedela jest prawdą niemożliwą do udowodnienia, w rzeczywistości, używając słowa, którego używa profesor Penrose, jeśli się nie mylę, jest „wiarą". Tyle że tutaj słowa „wiara" używam w szerszym znaczeniu i, jeśli pozwolicie, w bardziej radykalny. Bo jeśli weźmiemy pod uwagę to, co powiedzieliśmy wcześniej o niemożności wyrażenia przez naukę prawd

niezaprzeczalnych, to musimy powiedzieć, w tym szerokim znaczeniu, że nauka jest potężną formą wiary, zdolną nawet do przekształcenia świata niż wiara religijna. Modlitwa, czyli przymierze ze świętością i boskością, była niegdyś najwłaściwszym środkiem przemieniania świata. Co więcej, na sposób, w jaki Goedel wierzy, że potrafi odróżnić prawdy intuicyjne od prawd dających się udowodnić, wpływa ten brak, a mianowicie kwestia znaczenia tego zdecydowanie ważnego słowa, słowa „prawda". Jeszcze jedna obserwacja, jeśli mam czas. Inteligencja, którą chcemy zbudować, jest inteligencją kogoś innego. Nie chcemy budować własnej inteligencji. Ale od początku w ogóle nie można zaobserwować drugiej osoby. Nie wpadam w zachwyt nad Sartrem, ale jedną z ważnych rzeczy, które powiedział, jest to, że ta druga osoba, ta, którą nazywamy naszym bliźnim, to „piekło". I dlaczego? Bo drugi człowiek jest niezgłębiony! Co widzę w drugiej osobie? Widzę, że to zachowanie. Zachowanie, które jako takie nie jest inteligencją, ale jest zachowaniem, które maszyna może odtworzyć. Na zakończenie chciałbym wskazać, że szukamy inteligencji, czyli produktywności lub braku produktywności sztucznej inteligencji. Ale biorąc pod uwagę sposób, w jaki rozumiano człowieka na Zachodzie, to człowiek „jest", „jest" naturalna inteligencja, sztuczna inteligencja. Przypomnijmy sobie platońską definicję: produkcja jest przyczyną, która sprawia, że rzeczy przechodzą od niebytu do istnienia. Jest to organizacja środków do produkcji celów. Taką definicję człowieka podaje Zachód. Ale to jest także definicja maszyny! Tyle że na razie maszyny nie mają żadnego celu w zasięgu wzroku. Człowiek jest tą maszyną, która organizuje środki w celu wytwarzania celów, mając na uwadze właśnie obecność celów, idealną obecność; Bacht mówił o celach „ideell vorhanden", celem jest „ideell vorhanden". Technologia ta jest z pewnością przeznaczona do kontroli. Jednak jest ona skazana na dominację, jeśli weźmie pod uwagę zdolność filozofii dotarcia do sedna tych istotnych znaczeń. Można więc już powiedzieć, że człowiek naturalny jest maszyną i rzeczywiście świat jest maszyną. Jeśli Platon ponownie definiuje byt jako „dynamis", czyli władzę, a więc jako „poiesis", to świat jest już mechanizmem, w którym zorganizowane są środki w celu realizacji celów. Dlaczego człowiek traktuje poiezę jako coś oczywistego? Odpowiem i zamknę. Bo człowiek chce żyć. Powiedziałem wcześniej, że platońska definicja poiesis, którą wszyscy uważamy za oczywistą, nie jest obserwowalna. Ale dlaczego we wszystkich naszych przekonaniach tak uparcie trzymamy się tego, co nieobserwowalne? Ponieważ gdybyśmy nie wierzyli w naszą zdolność do tworzenia i niszczenia, do przenoszenia rzeczy z istnienia do nieistnienia i odwrotnie, nie żylibyśmy. A my chcemy żyć. Teraz chciałbym pozostawić Państwu taką możliwość. Czy my, ludzie, jesteśmy pewni, że mamy tylko „chęć życia"? I że nie ma w nas czegoś radykalnie bardziej zdecydowanego niż życie, czegoś bardziej istotnego niż życie, którego pragniemy, aby zapanować nad światem? Aby zdominować świat, musimy wiedzieć, jak go przekształcić. Dlatego musimy wierzyć w jego transformowalność. Ale czy na pewno jesteśmy po prostu „życiem"? A może inteligentne życie? Czy jesteśmy pewni, że za tym, w co wierzymy, że jesteśmy, nie kryje się coś zasadniczo bardziej zdecydowanego?

Penrose: Druga interwencja

Dziękuję za umożliwienie mi odpowiedzi na te bardzo interesujące uwagi i spostrzeżenia. Odniosę się tylko do kilku kwestii, które, jak sądzę, zrozumiałem. Jednym z nich była kwestia tego, co mikrotubule mają do powiedzenia na temat życia po śmierci: moim zdaniem nic! Nigdy nie wypowiadałem się na temat życia po śmierci, a już na pewno nie w tym kontekście, choć czasami robi to mój kolega Stuart Hameroff. Ale nie do końca zgadzam się z tym, co mówi na ten temat. Nie wiem. Nie mam nic do powiedzenia. Nie mam dowodów, aby wierzyć, że w jakimkolwiek sensie kontynuujemy życie po śmierci. Wydaje mi się mało prawdopodobne, aby coś takiego jak wspomnienia, które zgromadziliśmy w ciągu życia, miało jakąkolwiek szansę na przeżycie śmierci. Nie oznacza to, że w jakiejś formie nasza świadomość może nie odrodzić się na późniejszym etapie, w jakiejś formie. Nie mam takiego poglądu, można to sobie wyobrazić, to prawda, ale nie widzę na to żadnych dowodów. I nie rozumiem, dlaczego mikrotubule miałyby odgrywać w tym znaczącą rolę, ponieważ nie przetrwają śmierci bardziej niż neurony. Więc odłożę to na bok... Spójrzmy na inne punkty. Wystąpiły problemy dotyczące testowania

świadomości. Myślę, że o testowaniu można powiedzieć wiele rzeczy. Jednym z nich jest rzeczywiście to, czym zajmuje się mój kolega Stuart Hameroff, a mianowicie znieczulenie ogólne. I jest to bardzo jasny test, ponieważ znieczulenie ogólne, znieczulenie ogólne to środki, które bezpośrednio wpływają na świadomość, więc jeśli można odkryć, co robią w mózgu, jest to bardzo ważne w związku z jaką częścią mózgu lub jakimi aspektami mózg, biorą udział w świadomości. Tak więc przeprowadzono badania, szczególnie te przeprowadzone przez Stuarta Hameroffa i kilku jego kolegów w niedawnym artykule w Nature, w których rzeczywiście badają ten temat i wydaje się, że w grę wchodzą tu białka. To bardzo interesujące, że jeśli chodzi o środki do znieczulenia ogólnego, cóż, istnieje ich cała gama, w tym gaz paraliżujący, a także inne materiały, które wydają się nie mieć ze sobą żadnego związku chemicznego. Pytanie więc brzmi: co oni robią? To nie jest proces chemiczny. Zachodzi tu pewien proces fizyczny, który można spróbować zrozumieć. Ale to nie jest moja dziedzina. Badają, jaki wpływ mogą mieć środki znieczulające ogólne, i mikrotubule stają się mocnym kandydatem. Wydaje się, że mikrotubule są bardzo zaangażowane w działanie środków do znieczulenia ogólnego. Myślę więc, że jest to ważny rodzaj testu, który można przeprowadzić. Chciałbym się martwić innym rodzajem testu, ponieważ większość dyskusji, jakie toczymy, dotyczy mózgów, mówimy o neuronach itd., a ludzie zawsze rozmawiają ze mną o tym, co tu mamy [Penrose dotyka swojej głowy], czyli mózg. Mózg ma około 1011 neuronów, ale móżdżek ma porównywalną liczbę neuronów i o wiele, wiele więcej połączeń między neuronami, więc można powiedzieć, że móżdżek ma obiektywną przewagę w wykonywaniu obliczeń z neuronami i tak dalej, i powinien być w stanie zrobić znacznie więcej niż mózg. Niemniej jednak działanie móżdżku wydaje się całkowicie nieświadome. Tak się dzieje, jeśli jesteś dobrym pianistą lub tenisistą: kiedy uczysz się tych rzeczy, w dużej mierze świadomie je kontrolujesz, ale wtedy móżdżek uczy się od mózgu, co ma robić, i robi to znacznie lepiej. Weź pod uwagę każdego, kto idzie ulicą. No cóż, może na początku będziesz musiał nauczyć się poruszać mięśniami itd., ale potem wszystko zacznie być kontrolowane nieświadomie, a dzięki tej nieświadomej kontroli, gdy robisz rzeczy, które nie wymagają zrozumienia, móżdżek robi to znacznie skuteczniej. Zatem jest coś zupełnie innego w działaniu mózgu i móżdżku. I co to jest? Martwiłem się tym od dłuższego czasu. Słyszałem ostatnio Stuarta Hameroffa o niedawnych odkryciach, które wydają się obejmować fale aktywności przepływające tam i z powrotem przez mózg, a świadomość wydaje się pojawiać tylko na pewnym poziomie i to jest poziom, na którym znanych jest wiele tych komórek jako komórki piramidalne. Komórki piramidalne mają dużą liczbę mikrotubul, znacznie więcej niż inne komórki, i nie występują one w móżdżku. Zatem po raz pierwszy usłyszałem o czymś, w czym widać wyraźną różnicę między móżdżkiem a mózgiem i wydaje się, że w tym przypadku komórki piramidalne odgrywają ważną rolę. Myślę, że jest to bardzo ważny obszar badań, można go posunąć dalej i zrozumieć znacznie lepiej. Jest jeszcze jedna rzecz, która bardzo mnie intryguje, a która ma związek z często wypowiadanym twierdzeniem o świadomym chceniu. Istnieje wiele eksperymentów, których korzenie sięgają dawno temu Benjamina Libeta, gdzie wydaje się, że istnieją dowody, zanim zostanie podjęta świadoma decyzja, że w mózgu zachodzi jakaś aktywność i on w jakiś sposób już wie, na co się zdecydujesz . Niektórzy mówią: cóż, jest to dowód na to, że proces decyzyjny świadomości tak naprawdę nic nie robi, jest to po prostu coś, co ludzie nazywają „epifenomenem", który pojawia się samoistnie, ale tak naprawdę nie powoduje niczego. Podczas tego spotkania miałem pewne pomysły, ale wydaje się całkiem prawdopodobne, że istnieje pewne odesłanie wstecz (mówiłem o tym już wcześniej). Pogląd, który wydaje się wynikać ze schematu opracowanego przeze mnie i Stuarta Hameroffa, odgrywa pewną rolę. W jakiś sposób wygląda to tak, jakby akcja miała miejsce wcześniej, ale tak naprawdę tak nie jest. Świadomość rzeczywiście ma skuteczne działanie, ale ten inny eksperyment daje ci się zwieść. Nie chcę się w to teraz zagłębiać, bo to zaprowadziłoby nas za daleko. Druga kwestia dotycząca testów... Podałem ten przykład dotyczący pozycji w szachach. W rzeczywistości zostało to opracowane na podstawie wcześniejszego zestawu pozycji szachowych, które zostały zaprojektowane przez Davida Norwooda i Williama Hartstona. Myślę, że było to sto pozycji szachowych, problemów szachowych i

zostały one przekazane pewnej liczbie ludzi i pewnej liczbie komputerów, czyli programów. Połowa problemów została zaprojektowana tak, aby była łatwa dla komputerów i trudna dla ludzi. Te, które wymagają wielu obliczeń i obliczeń, ale nie było oczywistego powodu, dla którego nie rozumiałeś strategii. Dzięki czystym obliczeniom (i może krótkiemu kursowi obsługi komputera) możesz rozwiązać problemy. Ale druga połowa problemów była taka jak ta, którą podałem. Jest jeden ładny, z rzędem pionków przechodzących w poprzek lub z niekompletnym rzędem, a biały gracz ma zadecydować, jaki ruch wykonać, a ruch jest uzupełnieniem tej bariery pionkami, a nie wzięcie wieży, co jest oczywistą rzeczą, którą robi komputer. Wyraźnie widać różnicę: faktyczne zrozumienie tego, co robi bariera pionków, w przeciwieństwie do czystej kalkulacji, którą wykonuje komputer. Są to więc testy, które rzeczywiście pokazują różnicę pomiędzy świadomym myśleniem lub świadomym rozumieniem a czystym obliczaniem. A teraz o kwestii kreatywności. Wiele osób zapoznało się z książką Margaret Bowden, a ja napisałem o tym książkę i dużo z nią rozmawiałem na ten temat. Pytanie brzmi: czy można zbudować komputer, który będzie robił kreatywne rzeczy? Nigdy nie myślałem o tym jako o bardzo dobrym teście, szczególnie dotyczącym grafiki, ponieważ nigdy nie można stwierdzić, czy jest to naprawdę twórcze? Czy robi coś wspaniałego, co przekazuje jakieś nowe uczucia czy coś? I możesz włożyć wyczucie w pracę przy komputerze. Myślisz, że może to mieć jakiś sens czy coś... Bardzo trudno stwierdzić, czy było to twórcze, czy nie. Dlatego celowo nie mówię o kreatywności. Kiedy na przykład mówię o twierdzeniu Gödla... najwyraźniej sam Goedel był niezwykle kreatywny w tworzeniu tego twierdzenia, które stworzył. Ale nie podkreślam tego. Nie doceniam kreatywności, jaką niewątpliwie wykazał się Goedel w tworzeniu tego rezultatu. Podkreślam, że ktoś, kto jest po prostu dobrym matematykiem, potrafi śledzić argumentację. Aby więc podążać za argumentem, aby go zrozumieć, nie musisz być kreatywny, nie musisz być kimś, kto wyciąga rzeczy znikąd. Rozumienie jest czymś, dzięki czemu można śledzić to, co zrobił Goedel, można zrozumieć, co on zrobił, i można zrozumieć, dlaczego wynik jest prawdziwy, jeśli ufa się systemowi aksjomatów, czyli regułom, że podają jedynie prawdy. Zatem „zrozumienie" to coś, w czym widać różnicę. Wręcz przeciwnie, bardzo trudno dostrzec różnicę między kreatywnością a przypadkowym wytwarzaniem czegoś innego niż to, co robiono wcześniej. Czy ta praca jest naprawdę twórcza, czy nie? Trzeba mieć armię artystów, żeby zdecydować, czy uważają, że to naprawdę kreatywne... To bardzo trudne. Jestem więc pewien, że bycie kreatywnym jest czymś innym i myślę, że jest w nim pewna pomysłowość, która różni się od tego, co robią komputery, gdy tworzą przypadkowe rzeczy. Wierzę, że jest duża różnica, ale niezwykle trudno ją przetestować i zachować w tej kwestii obiektywizm. Kolejna poruszona kwestia dotyczyła Stephena Hawkinga. Z pewnością przedstawił kwestię, z którą się nie zgadzam, dotyczącą zagrożeń związanych ze sztuczną inteligencją. Stephen był zdania, że komputery będą robić rzeczy, które będą mądrzejsze od nas, lub powiedziałbym, inteligentniejsze od nas, gdy będą miały większe możliwości obliczeniowe. Po prostu nie zgadzam się z nim w tej kwestii. Myślę, że dzieje się coś zupełnie innego i o to mi chodziło. Myślę, że być może to, co mówiłeś o Stephenie Hawkingu, dotyczyło jego zdecydowanego ateizmu. Mówił, że poza nią nie ma nic, a świat powstał z niczego i do niczego zmierza. Sama nie podzielam takiego poglądu. Myślę, że coś tam jest. W ogóle nie nazwałbym tego poglądem religijnym. Powiedziałbym jednak, że w pewnym sensie istnieje tam jakieś znaczenie. Nie chciałbym przypisywać tego żadnemu poglądowi religijnemu, niemniej jednak uważam, że poglądy, które próbuję przedstawić, zawierają coś, co nadaje rzeczom znaczenie i że istnieje coś w świadomości, co nadaje temu znaczenie, i że jest na świecie coś, co ma wartość. Powiedziałbym też, i to dość zdecydowanie, że świadomość nie jest cechą zastrzeżoną tylko dla ludzkości i mocno wierzę, że jest to prawdą w przypadku zwierząt, na pewno jest to prawdą w przypadku psów, jest to prawdą w przypadku słoni, małp, goryli, delfinów... Wierzę, że wiewiórki również mogą mieć tę cechę. I myszy. Wiem, że myszy potrafią być w pewnym sensie niezwykle kreatywne... ponieważ napadły na nasz dom, a ja miałem takie pułapki, które nie zabijały myszy. Ale umiejętność myszy, które muszą wejść do środka, przejść przez pułapkę, oczyścić rzecz i uciec z kawałkiem czekolady... Mam ogromny szacunek do myszy. Myślę, że robią to ze świadomością,

myślę, że mają tę zdolność i myślę, że musimy szanować tę świadomość nie tylko u ludzi, ale także u innych stworzeń. Mam nadzieję, że to odpowie na Twoje komentarze.

Severino: Odpowiedz Penrose'owi

Jak zrozumiałem sens tego spotkania, miała to być wymiana poglądów pomiędzy naukowcami i filozofami. Powiedziałbym teraz (i przepraszam, jeśli mam wrażenie, że się przechwalam), że podjąłem pewne próby zrozumienia czegoś z nauki. Nie widzę odwrotnej próby ze strony naukowców, prof. Penrose'a. Rozważenie niektórych kwestii, które poruszyłem... Na przykład, mówiąc o „manifestacji świata", nie miałem na myśli cechy nieznanej kulturze dominującej. Mówi o tym Kartezjusz, mówi o tym Kant, w pewnym sensie mówi o tym Brouwer. Zamiast tego profesor Penrose pyta: „Gdzie leży świadomość?" Profesorze Penrose, usunął pan to, co powiedziałem i nie wziął pan tego pod uwagę. Ponieważ zapytać „gdzie świadomość jest częścią tej manifestacji, którą filozofowie nazywają transcendentalną, ale być może lepiej byłoby nazwać ją „nieprzekraczalną", tej manifestacji świata, która jest pierwotną formą świadomości. Dlaczego nie rozmawialiśmy na kolejny temat, o którym chciałbym usłyszeć dyskusję (również dlatego, że naukowcy zawsze interesowali się filozofią; myślę o Einsteinie, myślę o Schroedingerze, myślę o Hilbercie, myślę o Weylu, krótko mówiąc, to nie jest tak, że nauka i filozofia zawsze żyły na dwóch różnych górach oddzielonych od siebie otchłań), cóż, ciekawie byłoby omówić na przykład koncepcję kreatywności, którą odróżnił pan, profesorze Penrose, od koncepcji rezultatu. Czy dokonywałeś rozróżnienia pomiędzy pojęciem kreatywności a pojęciem, że tak powiem, produkcji? (To jest pytanie.) Tak, więc odróżniłeś pojęcie kreatywności od pojęcia produkcji. Następnie w swoim wykładzie wprowadziłem definicję przejścia od niebytu do bytu, definicję, która – jak zauważyłem – oznaczała, że nie byłaby ona obserwowalna. W tym momencie spodziewałbym się, że fizyk lub matematyk podskoczy z miejsca! Bo stwierdzenie, że pojęcie produkcji, przejścia od niebytu do bytu jest nieobserwowalne, oznacza, że nie można go eksperymentować i w pewnym sensie nie ma ono charakteru naukowego! To jest teoria. Mam na myśli na przykład, że dzisiaj widzimy wokół siebie kolory, ale nie widzimy świadomości innych obecnych ludzi. Tutaj także spodziewałem się poruszyć kwestię profesora Penrose'a. Świadomość innych nie tylko nie należy do myszy i wszystkich zwierząt, o których wspomniałeś, ale nawet nie widzimy jej w tym, co nazywamy „inną osobą". A może dostrzegasz inteligencję tych, którzy stoją przed tobą? Świadomość tych, którzy są przed tobą? Jeśli przez świadomość rozumiemy pojawienie się lub manifestację świata... Inną kwestią, która byłaby interesująca i powiązana z tym, jest wszechobecność przejścia od niebytu do bytu, co, jak powiedzieliśmy, nie jest obserwowalne ani eksperymentalne . To jest obecne wszędzie. Nawet tutaj możemy wrócić jeszcze raz do Gödela, który stworzył (odkryto to po jego śmierci) dowód na istnienie Boga w stylu argumentacji ontologicznej, w wersji leibnizowskiej. Goedel wypisuje dowód na istnienie Boga. Co to znaczy? Bóg dla Goedla jest bytem koniecznym. Oznacza to, że jeśli chodzi o punkt wyjścia jego dowodu, punktem wyjścia jest zbiór bytów niepotrzebnych. Mianowicie zbiór bytów, które są i nie są, które powstają z niebytu i zaczynają istnieć. Zatem także, powiedziałbym nawet, Goedel uczestniczy w tej platońskiej koncepcji „poiesis", której dominację stwierdzam, dominację tym większą, im więcej jest znane po stronie wiedzy naukowej, ale dzisiaj powiedziałbym Niestety, także po stronie wiedzy filozoficznej. To tylko niektóre z rozważań, o których chciałem wspomnieć. Ale przychodzi mi do głowy jeszcze jeden. Wiem, jakie znaczenie profesor Penrose przypisuje słowu „wiara". Tutaj też powiedziałem, że nauka jest wiarą w jeszcze szerszym znaczeniu. I do cholery, każdy naukowiec powinien uderzyć pięścią w stół i powiedzieć „tak" lub „nie"! Jeśli jednak powie „nie", powinien podać dobry powód. Jeśli chodzi o tę delikatną kwestię dotyczącą praktycznych aspektów nauki, cóż, jeśli nie będziemy o tym rozmawiać, to o czym będziemy rozmawiać, gdy filozofowie spotkają się z fizykami? ...Praktyczny aspekt nauki. Oznacza to, że pojęciowa artykulacja wiedzy naukowej pozwala na władzę nad światem przewyższającą inne pojęciowe artykulacje, takie jak pojęciowa artykulacja przymierza z sacrum, jak powiedziałem wcześniej, czy modlitwa. Oczywiście

konceptualna artykulacja wiedzy naukowej robi wrażenie! To właśnie pozwala dziś na największą władzę nad światem. Ale prawda to jedno, a władza to drugie. Jeśli chodzi o koncepcję władzy, przychodzi mi na myśl inny temat. Mówimy o władzy nad rzeczami, ale wspomnijmy Karla Poppera, który powiedział: Gdyby Robinson Crusoe wynalazł całą współczesną naukę na bezludnej wyspie, wynalazek ten nie miałby wartości naukowej. Dlaczego? Bo nie zostałoby to intersubiektywnie rozpoznane. Aby była moc, nie wystarczy, że bomba atomowa wybuchnie. Musimy zdać sobie sprawę, że to pęka. Bo inaczej jest bezsilny. Jak mówi Karl Popper (kolejny autor, nad którym nie wpadałbym w zachwyt, ale chociaż z tego powodu pamiętajmy o nim): aby istniała władza, potrzebne jest intersubiektywne uznanie. Ale co oznacza uznanie intersubiektywne? Oznacza to zauważenie, że inni dostrzegają przemianę świata. Ale czy to obserwacja, czy interpretacja? Czy rozpoznajemy świadomość innych? NIE! Co wiemy o zachowaniach językowych innych? Wiemy jedno: że jest to wydarzenie, które interpretujemy w określony sposób. Nie ma w tym nic obiektywnego. Zatem samo istnienie mocy (i to powinno być czymś, co wysadza fizyków w powietrze...) jest czymś interpretowanym, a nie czymś obiektywnym. To jest coś pożądanego. Chcę, żeby otaczający mnie ludzie uznali tę moc. Mógłbym wymyślić wiele innych rzeczy do powiedzenia, ale nie chcę nadużywać cierpliwości obecnych.

Penrose: Odpowiedź Severino

Dziękuję za te komentarze. Mam nadzieję, że mi się uda... Muszę przyznać, że tak naprawdę nie rozumiem niektórych kwestii, które poruszyłeś. Pozwólcie, że powiem, jak postrzegam sprawy, które mnie interesują i o których mówiłem. Martwię się... Mój sposób patrzenia na naukę polega na próbie ustalenia prawdy o świecie. Zatem nie ma tu żadnego problemu moralnego. To znaczy, to osobne pytanie. Próbujemy dowiedzieć się jaka jest prawda o świecie. Jak działa świat. Czy to działa w ten sposób? Czy to działa w ten sposób? Nauczyliśmy się czegoś, wracając do Newtona, i dowiedzieliśmy się, że niektóre z tych rzeczy są bardzo głębokie, głębokie i prawdziwe, a potem dowiedzieliśmy się, że inne nie są do końca prawdziwe, a potem pojawiły się inne rzeczy, które udoskonaliły nasze zrozumienie i teraz mamy znacznie lepsze zrozumienie rzeczy i wiemy, jak działa świat, z dużą precyzją. Niekoniecznie tylko precyzja, ale to poziom głębszy niż dotychczas. A ja postrzegam naukę jako próbę dowiedzenia się, co dzieje się na świecie, co nim steruje. Osobną kwestią jest to, jak wykorzystać naukę. To jest technologia. Istnieje technologia, która wykorzystuje naukę i tworzy rzeczy, i z pewnością ma wpływ na to, co robią ludzie, i jest to rzecz jasna ważna. Można zobaczyć, jak nauka wpływa na życie ludzi w istotny sposób, nie zawsze dobry, czasem zły, a są to rzeczy, które często wynikają z naukowego zrozumienia. Ale nie tym zajmuje się nauka. Nauka próbuje zrozumieć sposób, w jaki działa świat. I tak właśnie postrzegam naukę. Jest jeszcze technologia, która ma związek z nauką, ale nie jest nauką. Ma to także związek z moralnością. Jest moralność. Możesz zrobić dobry użytek z nauki lub możesz zrobić z niej zły użytek. Albo dobre wykorzystanie technologii, albo złe wykorzystanie technologii. To osobne pytanie. Myślę jednak, że bardzo ważne jest, aby poznać prawdę na ten temat, a wtedy znacznie jaśniej będzie można powiedzieć, co jest korzystne, a co nie. Myślę jednak, że w tej chwili jest to bardzo niejasne. Na przykład rozwój komputerów, sposób ich działania i sposób działania. A ludzie uważają, że są bardzo inteligentni, choć ja tak nie uważam. Cóż, mój ojciec miał kiedyś maszynę Brunsviga, odwracasz rękę i ona wykonuje obliczenia, które są o wiele bardziej skomplikowane, niż jest w stanie zrobić człowiek. OK, ale widać, że to urządzenie mechaniczne i robi rzeczy, których człowiek nie jest w stanie zrobić bardzo łatwo lub może zrobić znacznie wolniej za pomocą kartki papieru i ołówka. Ale to nie jest nic, co wiąże się z inteligencją. To nie wymaga inteligencji. I myślę, że komputery, które mamy dzisiaj, naprawdę nadal takie są. Nie wymagają inteligencji, nie wymagają rozumienia świata. I ważne jest, aby o tym wiedzieć. Zatem punkty, które próbuję poruszyć, dotyczą prawdy o świecie, a przynajmniej do tego dążę. A teraz, czy jest to korzystne dla ludzkości? Nie wiem! Ponieważ na przykład możesz wykorzystać tę technologię do rzeczy, którymi bardzo się martwię. Można opracować maszyny, takie jak drony, które mogą wysadzić kogoś w powietrze z dużej odległości, prowadzone przez osobę

siedzącą w środku Teksasu przed ekranem komputera. Czy to jest moralne? Cóż, nie sądzę, żeby to było moralne. Rodzi to jednak pewne pytania, których wcześniej nie było. Z pewnością więc technologia wprowadza pytania moralne. Nie porusza kwestii moralnych. Kiedy więc mówię, że zajmuję się nauką, próbuję zrozumieć sposób, w jaki działa świat i nie szukam władzy (myślę, że coś w tej kwestii ująłeś). Nie próbuję kontrolować świata. Nie na tym polega nauka. Na tym właśnie polega technologia. Dotyczy to także dobra i zła, więc pojawia się moralność. To są odrębne kwestie, zależą od nauki, jeśli nauka się zmieni, to będą dotyczyć innych dziedzin, technologii i moralności, trzeba szanować naukę i widzieć co im to mówi. Jednak w Waszych komentarzach nie byłem do końca pewien, czy krytycznie odnosicie się do tego, że nie odniosłem się do niektórych z tych kwestii, ponieważ myślałem, że te kwestie nie wchodzą w zakres moich zadań. Myślę, że na kwestię moralności wpływa to, jak daleko możemy odnaleźć świadomość w królestwie zwierząt. A ja nie znam odpowiedzi na to pytanie. Z pewnością uważam, że świadomość nie ogranicza się do istot ludzkich i jeśli to zrozumiemy, a mianowicie, że świadomość obejmuje inne zwierzęta, wówczas nasza moralność jest bardzo ważna w stosunku do innych zwierząt. Powiedziałbym więc, że ma to duży wpływ na moralność, ale nie o tym głównie mówiłem. Myślę, że najpierw musimy zrozumieć te rzeczy, a potem może zobaczymy, jaki to będzie miało wpływ na inne kwestie. To wszystko, co mogę powiedzieć. Obawiam się... Mam nadzieję, że to pomoże w tym, co mówisz.

Śmierć umysłu cesarza z perspektywy eternalistycznej

Wstęp

W 2017 roku niektóre ważne gazety podały jako sensacyjną wiadomość fakt, że dwa chatboty przeszkolone przez grupę Facebook Artificial Intelligence Research (FAIR) na korpusie rozmów tekstowych w języku angielskim z udziałem piłek, czapek i książek rozpoczęły autonomicznie rozmawiać ze sobą, używając języka angielskiego w sposób inny niż ludzki. Ponieważ badacze nie byli w stanie zrozumieć ani przerwać dziwnego dialogu między dwoma chatbotami, postanowili „zabić" dialogistów, wyłączając ich. Pomimo tego, że w dziedzinie sztucznej inteligencji (AI) uczeni twierdzą, że możliwe jest wywołanie ewolucji języka w systemach wieloagentowych, które przy odpowiednich umiejętnościach mogą współpracować przy zadaniu, a następnie wymieniać zestaw symboli które służą jako tokeny w wygenerowanym języku, problem wydaje się wiązać z pewnymi trudnymi pytaniami. Rzeczywiście, mimo że zakłada się, że systemy AI (AIS) mogą generować nowy język na podstawie początkowej serii terminów i reguł, trudność polega na rozważeniu, co oznacza autonomia mówienia, gdy nie ma dwujęzycznych tłumaczy, którzy mogliby wyjaśnić treść tego języka nowy język. Problem w tym, że ludzka inteligencja może wnioskować i w ten sposób tłumaczyć innowacyjne treści generowane przez AIS jedynie wtedy, gdy problem zostanie rozwiązany w pożądany sposób. Ponieważ w obliczeniowym procesie informacyjnym można zrozumieć początek i koniec procesu, pośrednia heurystyka staje się możliwa do przełożenia poznawczego i językowego. Jeśli jednak ścieżka zejdzie z drogi, zakładając inne skupienie, które nie będzie skierowane na zrozumiałe rozwiązanie – jak w przypadku dwóch zabitych chatbotów – ludzkiej inteligencji trudno będzie wywnioskować, o czym myślą te AIS, a potem mniej więcej mniej celowa komunikacja. Zadanie obliczeniowe podjęte przez dwa chatboty mogło wywołać proces stochastyczny, uniemożliwiając sterowanie maszynami. W cybernetyce procesy komunikacji wymagają zbudowania języka dwujęzycznego przydatnego zarówno dla człowieka, jak i sztucznej inteligencji. Podważa to jednak naturalny, neuronaukowy pogląd na język, wyrażany na przykład przez neurologa Terence'a Hinesa, który twierdzi, że tylko ludzie komunikują się za pomocą języka. Rzeczywiście, po opisaniu specjalizacji ludzkiego mózgu w przetwarzaniu języka i pokazaniu, jak ta specjalizacja manifestuje się na poziomie funkcjonalnym, anatomicznym i komórkowym (a wszystko po to, aby potwierdzić, że świadomość jest materią szarą), Hines stwierdza, że „Język jest wyjątkowy dla istoty ludzkie. Inne gatunki mają swoje własne, często złożone i piękne systemy komunikacji, ale

tylko ludzie mają język” . W dziedzinie sztucznej inteligencji twierdzenie to wydaje się oczywiście absurdalne, ponieważ komunikacja językowa jest podstawą wszelkich badań i praktyk cybernetycznych, dlatego też trudno twierdzić, że komputery nie komunikują się z człowiekiem. Zamiast tego prawdziwy problem może dotyczyć empatii lub możliwości, że język używany do komunikacji pozwala na wyciąganie wniosków na temat świadomości/umysłu mówiącego. Języki naturalne to media umożliwiające wnioskowany dostęp do świadomości/umysłów innych ludzi, ale co z językami, które pośredniczą w relacjach między ludźmi a sztuczną inteligencją? Właściwie problem ten pojawił się po raz pierwszy w odpowiedzi na test Turinga. Twierdzenie Turinga stwierdzało, że ludzki intelekt, choć czasami potężniejszy od komputerów, nie jest zasadniczo lepszy [66]. Oznacza to, że człowiek może pomylić doskonale zaprogramowany AIS z człowiekiem prowadzącym ślepy dialog i właśnie to mierzy test Turinga. Silna perspektywa sztucznej inteligencji, która w dużej mierze jest zgodna z psychologią poznawczą – czyli tak zwanym „przetwarzaniem informacji przez człowieka” [35] – kładzie nacisk na istnienie strukturalnych i funkcjonalnych powiązań między ludzkimi umysłami a komputerami. Jeśli tak, można postawić hipotezę, że oprogramowanie może nauczyć się dużo tak samo, jak dziecko, ostatecznie będąc w stanie rozwiązać każdy problem, używając odpowiedniego języka, bez ludzkich wskazówek wyrażanych w językach naturalnych [50]. Kolejnym logicznym pytaniem jest, czy w procesie rozwojowym AIS może powstać intencjonalny aparat poznawczy, np. skupiony na celach, którymi można zarządzać autonomicznie i bez kontroli człowieka. Rzeczywiście, jeśli AIS będzie w stanie działać lepiej niż ludzka inteligencja i stanie się autonomiczny, jego cele mogą być niebezpieczne dla życia ludzkiego. Jaki rodzaj świadomości możemy zatem wywnioskować z AIS, które uniezależniają się od nas i naszego doświadczenia? Rozwiązanie kryje się w starej kwestii świadomości, wyrażonej w pytaniu o umysł – „materia” – tożsamość, niezależnie od tego, czy materia jest postrzegana jako szara (mózg), czy nie. Roger Penrose nie zgadza się z silną tezą dotyczącą AI. Jego zdaniem AIS działa wyłącznie poprzez język formalny, podczas gdy myśl człowieka charakteryzuje się funkcjonowaniem, którego nie można sprowadzić do procesów obliczeniowych – jak na przykład aha-Erlebnis (doświadczenie aha czy efekt eureka). Kierując się tą logiką, jeszcze trudniej jest wywnioskować, czy istnieje jakikolwiek poziom świadomości podobny do ludzkiego w autonomicznie komunikującym się AIS. Ponieważ, jak twierdzi Penrose, znaczeniem każdego dyskursu formalnego jest treść rzeczywistości, problemem jest zrozumienie, jaka jest treść tego dyskursu. Jeśli, jak twierdzi Penrose, ich dyskurs ma charakter wyłącznie formalny, to jest mniej istotny niż myśl ludzka, która zawiera odniesienia do konkretnej rzeczywistości. Jeśli jednak ich dyskurs jest czymś więcej niż formalnym, a ludzie po prostu nie znają jego treści odnoszących się do rzeczywistości, to jak ludzie mogą rozpoznać tę treść, skoro nie rozumieją dialogu? W tym przypadku „rzeczywistością”, do której ludzie nie mają dostępu, są mentalne lub materialne stany AIS, które są konkretnie rzeczywiste, ale funkcjonują w stanie, który nie jest do końca zrozumiały, w związku z czym nie można wywnioskować, co dzieje się w ich komunikację. Wyglądało na to, że chatboty reagowały na siebie nawzajem, jakby celowo odnosiły się do czegoś, czego obaj byli świadomi. Problem ten jest istotny, gdyż dotyczy relacji myśli do rzeczywistości. Z tej dziedziny wypływają pozornie nieskończone argumenty, które nieuchronnie przyciągają powszechną uwagę, a stanie się to jeszcze bardziej prawdziwe w miarę ulepszania sztucznej inteligencji, tworząc nowe reprezentacje niematerialnego umysłu ze względną niezależnością od materii. Rzeczywiście, ten sposób myślenia jest tak interesujący, ponieważ wywołuje znakomicie metafizyczne zagadki, nawet jeśli pozornie dąży do celów utylitarnych, takich jak tworzenie nowych technologii w celu poprawy jakości życia ludzi. W szczególności przywołuje takie pojęcia jak dusza, duch, umysł i świadomość oraz pyta o relacje pomiędzy tymi pojęciami a materią (szarą lub inną); na przykład, czy jedno można zredukować do drugiego? Podstawowym pytaniem jest, czy umysł/świadomość może w jakiś sposób to zrobić aby przetrwać zakłócenie jej materialnego wsparcia, a następnie czy jest w stanie przenieść się do innego medium, aby przetrwać. Niniejszy artykuł odpowiada na te pytania, czerpiąc z wiecznej perspektywy Emanuele Severino i badając implikacje

świadomości kwantowej Penrose'a dla życia pozagrobowego, czyli śmiertelności umysłu Imperatora. Rozumowanie będzie przebiegać przez kilka etapów: najpierw przedstawiona zostanie najnowsza historia koncepcji „duszy"; następnie wyjaśni problem tożsamości umysłu, materii i materii (materia jest szara lub inna) w odniesieniu do kwestii świadomości AI; następnie rozważy możliwość przetrwania świadomości po zniszczeniu jej materialnego wsparcia; wreszcie rozważona zostanie aktualność tych tematów dla zagadnień związanych ze śmiercią, związanych z ontologicznym pytaniem prawdy.

Tradycyjne reprezentacje duszy w życiu pozagrobowym i ich znaczenie psychologiczne

Współczesną myśl zachodnią określa się jako świecką i postmodernistyczną, ponieważ dominuje w niej logika naukowa i technologiczna. Jest to wynik radykalnego obalenia metafizyki, a co za tym idzie religii, która na metafizyce opierała się – a co za tym idzie i swoją wiarygodność. Oznacza to, że problem śmierci, którym zajmowały się wcześniej religie historyczne, wszedł w swego rodzaju kryzys. Tam, gdzie kiedyś metafizyczne i religijne koncepcje duszy i świadomości zakładały esencję, która pozostaje po śmierci, nawet po śmierci materialnego ciała, współczesna kultura zachodnia odrzuca koncepcję niezniszczalnej esencji, która trwa w zaświatach . Najlepiej ugruntowane argumenty dotyczące reprezentacji śmierci zawarte są w koncepcjach binarnych „dualizmu kontra monizmu". Dualistyczne teorie metafizyczne leżą u podstaw tradycyjnych religii i konfigurują duszę (lub świadomość) jako zasadę tożsamości jednostki, której przeznaczeniem jest przetrwać w zaświatach. Ta zasada tożsamości jest wyposażona w samoświadomość, pamięć i intencjonalność/sprawczość, a niektóre zachodnie teorie stwierdzają, że ta część istoty ludzkiej trwa po śmierci w nieskończonej dziedzinie nagród i kar (niebo, piekło itp.), zakładając, że „silna" perspektywa dualistyczna, która jest zasadniczo platońska i kartezjańska . Słaba forma perspektywy dualistycznej zakłada trwałość tylko części tej esencji po śmierci, przy czym część, która pozostaje, staje się odsubiektywizowaną i zubożoną zasadą tożsamości. Według różnych przedstawień życia pozagrobowego zasada tożsamości może zostać ograniczona w następujący sposób: życie pomniejszone (np. grecki Hades, szkoła żydowska czy kraina umarłych); niewidzialna egzystencja ziemska (kultury przedontologiczne i mityczne); oddzielenie ciała od duszy, a następnie metempsychoza, utrata samoświadomości lub początkowej tożsamości i ostateczne połączenie z całością (hinduizm, buddyzm, orfizm i neoplatonizm, gnostycyzm). Między XIX a XX wiekiem w krajach Zachodu religie straciły wiarygodność, gdy technonauka zastąpiła praktyki religijne praktykami odpowiadającymi materialistycznym potrzebom jednostek. Od tego czasu wiele zasad moralnych i politycznych uległo zmianie na korzyść demokracji i dobrobytu ludzi. Kiedy na scenie politycznej dominowały zachodnie religie metafizyczne, prześladowania i wojny w imię Boga były powszechne, pogłębiając codzienną walkę z biedą, chorobami i zmęczeniem. Mimo to religie Zachodu przez wieki potrafiły utrzymać swą kulturową dominację, orientując postawy i zachowania społeczne aż do męczeństwa. Według pozytywizmu i dialektyki materialistycznej siła perswazji religii wiązała się z tym, że niosły one pocieszenie w obliczu ludzkiego cierpienia i nadawały sens śmierci . Perspektywa poświęcenia życia dla Boga w celu uzyskania wiecznego szczęścia po śmierci pomagała ludziom znosić ból życia w społeczeństwie pełnym przemocy i przerażający strach przed śmiercią, pomimo udziału religii w budowie tego nieliberalnego i pełnego przemocy społeczeństwa. W społeczeństwie świeckim, które jest bardziej liberalne i pluralistyczne, wiarę w życie pozagrobowe definiuje się jako „myślenie życzeniowe". Mimo to teoria zarządzania terroryzmem (TMT) w przekonujący sposób wyjaśnia, dlaczego ta wiara jest tak istotna psychologicznie, pokazując, w jaki sposób wyjątkowa świadomość śmierci u ludzi powoduje potencjalnie wyniszczający terror egzystencjalny. Terror ten jest zarządzany poprzez przyjęcie kulturowych światopoglądów, które nadają światu znaczenie i nadają poczucie indywidualnej wartości, czyniąc w ten sposób jednostkę kwalifikującą się do dosłownej i/lub symbolicznej nieśmiertelności [4, 58]. Kiedy śmiertelność staje się szczególnie widoczna, aktywowane są dwa poziomy obrony: „obrona bliższa", która utrzymuje wyraźną świadomość śmiertelności poza świadomością ogniskową poprzez zaprzeczanie, rozproszenie lub racjonalne zachowanie

instrumentalne, oraz „obrona odległa", która wzmacnia wiarę w światopogląd kulturowy. co podnosi samoocenę i poprawia relacje z bliskimi osobami. Ponad 1000 badań TMT pokazuje wszechobecny wpływ „przypomnień o śmierci" na postawy, uczucia i zachowania ludzi, sprzyjając uprzedzeniu w grupie oraz wzmacniając uprzedzenia i dyskryminację wobec grup obcych. Teza ta jest szeroko poparta dowodami i była filozoficznie antycypowana przez oświecenie, pozytywizm, neopozytywizm i materialistyczny redukcjonizm, które teoretycznie rozważały, w jaki sposób jakiekolwiek odniesienie do Boga jest niewiarygodną hipostatyzacją ludzkiego pragnienia istnienia po śmierci. Radykalne odrzucenie jakiejkolwiek metafizycznej, absolutnej zasady, której można przypisać spowodowanie świata, sugerowało również, że fizyka nie opiera się na żadnej transcendentalnej zasadzie w celu wyjaśnienia funkcjonowania wszechświata i wszystkich jego substancji materialnych. Dlatego też w dzisiejszych czasach ludzie Zachodu wydają się być szczególnie narażeni na udrękę śmierci, nieustannie starając się posiadać formę istnienia poza śmiercią. Jednym z przykładów tego jest technologiczne życie pozagrobowe, w którym autonomiczna, inteligentna i samoświadoma zasada tożsamości przenika pomiędzy materialnymi nośnikami.

Solipsyzm i problem relacyjny/tożsamości umysł–mózg

Aby epistemologicznie rozwiązać problem nieodłącznie związany ze zdolnością człowieka do rozpoznawania form autonomicznego umysłu/świadomości w AIS, potrzebujemy dwóch kolejnych kroków: po pierwsze, musimy uporać się z problemem solipsyzmu, aby w znaczący sposób wyobrazić sobie formę intersubiektywności człowiek-komputer, która wymaga, jako drugiego kroku, zdefiniowania związku pomiędzy umysłem/świadomością a materią fizyczną. Koncepcja solipsyzmu opiera się na założeniu, że jedyną rzeczą, którą jednostka może stwierdzić z całą pewnością, jest jej własne istnienie. Wprowadza to możliwość zwątpienia we wszystko, ponieważ wszystko, co jednostka postrzega, może być iluzją. Co więcej, jednostka nie może widzieć świadomości innych ludzi, a każda empatyczna reprezentacja jest zawsze stronnicza ze względu na jej/jego osobę, jej osobistych potrzeb i poglądów. Metafizyka rozwiązała ten problem postulując istnienie Boga, pojmowanego jako pierwsza i ostatnia zasada przyczynowa (Arystoteles) – bytu złożonego z istot radykalnie odmiennych od materii i obdarzonego mocą uzasadniania materii (Platon) – gwarantującego wiarygodność postrzeganych rzeczywistość (Kartezjusz). Nasze niematerialne esencje, które rzekomo mają tę samą naturę co Boża i powracają do Boga po śmierci, są uniwersalne i stanowią haecceitas – czyli „tę" – rzeczy (Albertus Magnus). Jako taka świadomość każdego człowieka jest rozpoznawalna na podstawie tej uniwersalnej istoty (duszy/ducha), która ma dążyć do najwyższego dobra: woli Bożej (Tomasz z Akwinu). Po obaleniu jakiegokolwiek wymiaru metafizycznego, współczesne teorie materialistyczne (monistyczne), zakłada, że żaden „meta" wymiar nie jest rzeczywisty – to znaczy, że poza rzeczywistością fizyczną nie istnieje nic, ani Bóg, ani esencje. Uniwersaliści metafizyki wymieniali duszę jako siedzibę świadomości, którą uważano za uniwersalnie identyczną dla wszystkich, którzy zostali nią obdarzeni, ponieważ była w stanie uczestniczyć w uniwersalnych, boskich ideach. Wraz z upadkiem tej perspektywy i późniejszym zastąpieniem koncepcji duszy koncepcją umysłu, nie było już możliwe wnioskowanie o świadomości innych. Jakakolwiek kluczowa, niewątpliwa koncepcja zwornikowa zdolna do utrzymania struktury pojęciowej niezbędnej do rozwiązania problemu solipsyzmu została zniszczona. Co więcej, po wyeliminowaniu jakiegokolwiek wymiaru uniwersalnego (transcendentalnego), problem tożsamości i samoświadomości w życiu i życiu pozagrobowym stał się nonsensowny, jak sugeruje definicja „myślenia życzeniowego":

Wiele osób pragnie przeżyć śmierć cielesną. Ale pragnienia tego nie sprawią. Myślenie życzeniowe jest zrozumiałą ludzką skłonnością, która jednak nie odgrywa żadnej roli w określaniu, jak się sprawy mają. Ilekroć napotykamy przypuszczenia na temat życia pozagrobowego, które są całkowicie pozbawione

uzasadnienia i niepoparte dowodami empirycznymi, odrzucamy je jako myślenie życzeniowe. Brak zasady myślenia życzeniowego: przypuszczenia motywowane myśleniem pobożnym są wykluczone.

Po obaleniu metafizyki, która miała na celu dedukcyjne i bezdyskusyjne rozwiązanie związku między prawami uniwersalnymi a konkretnym, fizycznym światem, problem solipsyzmu wydawał się nie do rozwiązania. Indukcyjne potwierdzenie istnienia i zawartości innych umysłów lub określenie, gdzie umysł jako substancja znajduje się w wymiarze czasoprzestrzennym, stało się niezwykle trudne – wręcz niemożliwe. Główna trudność polegała na zdefiniowaniu umysłu, w tym świadomości, w odniesieniu do materii fizycznej, która go wspiera – mózgu lub ewentualnie komputera. W istocie zasadę „braku myślenia życzeniowego" propagował głównie materialistyczny redukcjonizm, który miał na celu ujednolicenie wiedzy i nadanie jej sensu realności poprzez powszechne przyjęcie języka nauk przyrodniczych. Z tego punktu widzenia możliwe jest skonfigurowanie intersubiektywności za pomocą języka fizykalizmu. Wśród tych, którzy uznali realistyczny potencjał solipsyzmu, byli Ludwig Wittgenstein i Rudolf Carnap. W swoim Tractatus logico-philosophicus (1922) Wittgenstein przenosi solipsyzm na obszar językoznawstwa; z tego fragmentu wynika pogląd, że granice „mojego świata" (świata, który „objawia mi się") są określone przez granice „mojego języka" (języka, który tylko ja rozumiem). W tej perspektywie pojęcie solipsyzmu jest „ścisłe": manifestuje się, a mimo to pozostaje niewypowiedziane, dlatego należy je przełożyć na stwierdzenie: „Jestem moim światem". Carnap [6] w swojej pracy The Logical Structure of the World przekształcił tę samą wersję pojęcia w „solipizm metodologiczny", odnosząc się do problemu wiedzy i podstawowych elementów, które pozwalają realistycznie zrekonstruować ogólną strukturę rzeczywistości poprzez język fizyki. Elementy te są tym, co Tractatus identyfikuje jako odpowiedniki pomiędzy każdym terminem językowym a tym, co rzeczywiste, i wynikają z wyboru traktowania ich jako bezpośrednich faktów doświadczenia. Podstawowym elementem teoretycznym, który jednoczy Wittgensteina i Carnapa w odniesieniu do solipsyzmu, jest koncepcja podmiotowości bez podmiotu, w której solipsyzm pokrywa się z czystym realizmem, ponieważ – jak wskazuje Carnap – „ja" nie jest absolutnie niczym oryginalnym. W tym sensie kartezjańskie cogito nie jest już uznawane za podstawę rozstrzygania hiperbolicznych wątpliwości i ogranicza się do bycia doświadczeniem: solipsystycznym doświadczeniem myślenia. Z pewnością projekt Koła Wiedeńskiego polegał na ustanowieniu jednoznacznego języka fizykalizmu: cała rzeczywistość jest materią, dlatego wiedza intersubiektywna jest realistyczna i musi być fizyczna. W tym języku, składającym się z tego, co maksymalnie „można powiedzieć", każde twierdzenie może od razu uchodzić za intersubiektywne. Oznacza to, że intencjonalność przekracza solipsystyczne doświadczenie refleksji dzięki wspólnej rzeczywistości fizycznej, która staje się intersubiektywnym przedmiotem myśli. Jerry Fodor twierdzi jednak, że koncepcja solipsyzmu metodologicznego Carnapa stoi w sprzeczności z absolutnym fizykalistycznym redukcjonizmem, który uniemożliwiałby wyjaśnienie funkcjonowania umysłu. Solipsyzm metodologiczny poznawczy zakładany przez psychologów można podsumować następująco: psychologia bada wewnętrzne stany ludzi lub organizmów tylko o tyle, o ile dają się opisać; składnia stanów psychicznych jest niezależna od procesów fizjologicznych mózgu; prawa psychologiczne odnoszą się tylko do stanów psychicznych. Stanowisko to promuje psychologię obliczeniową, która nie uznaje przejrzystości ani dostępności stanów mentalnych (solipsyzm), ale identyfikuje pozytywne modalności obserwacyjne między zachowaniem, językiem i reprezentacją stanów mentalnych. W ten sposób można wnioskować o stanie psychicznym podmiotu, nie przypisując jednak temu wnioskowi statusu niewątpliwości (tj. w odniesieniu do umysłu znamy tylko to, co poznawalne, czyli właśnie to, co obliczalne). W swojej charakterystyce obliczeniowej psychologia pozytywistyczna uważa procesy poznawcze zarówno za symboliczne (reprezentacyjne), jak i formalne (syntaktyczne), a także za należące do obszaru umysłu, którego nie da się całkowicie sprowadzić do istoty szarej (mózgu). Jednakże ten dualizm nie postuluje żadnej nieśmiertelnej esencji w umyśle; potwierdza po prostu, że umysł i mózg to dwie różne istoty. Różne odmiany tego dualizmu stwierdzają,

że między tymi bytami istnieją różne relacje lub nie ma ich wcale, w zależności od aspektów branych pod uwagę przez różne nurty badawcze oraz w zależności od tego, czy dana odmiana dualizmu odnosi się do substancji (umysł jest byt niefizyczny oddzielony od ciała) lub właściwości (nie ma bytu niefizycznego, lecz świat fizyczny posiada pewne właściwości należące do odrębnej klasy) . Na przykład interakcjonizm uważa, że umysł jest nieredukowalny do mózgu, jednocześnie dopuszczając związki przyczynowe między nimi, funkcjonalizm stwierdza, że świadome procesy umysłowe są formą adaptacji biologicznej i należą do fizycznych systemów opracowania, które nie są wyłącznie ludźmi , Emergentyzm podkreśla twórczy, systemowy aspekt zmian ewolucyjnych w procesie adaptacji , Sperry, korelacjonizm zakłada związek pomiędzy umysłem przeciwieństwa mózgu itd. Podsumowując, możemy powiedzieć, że istnieje istotna różnica pomiędzy metafizycznym dualizmem substancji a niemetafizycznym dualizmem własności: w przeciwieństwie do pierwszego, ta ostatnia wyklucza jakąkolwiek niezmienną, niematerialną istotę i obejmuje wszystkie rozróżnienia między fizykalizmem redukcyjnym a funkcjonalizmem fizykalistycznym, Emergentyzmem, interakcjonizmem, korelacją itd., zakładając, że wszystkie substancje są fizyczne, ale stany mentalne są powiązane ze stanami mózgu za pomocą wielu form właściwości . Założenie to opiera się na dowodach pokazujących, w jaki sposób funkcje umysłowe zachodzą w mózgu. Przykładowo neurolog kliniczny David Weisman ([68], s. 84) opisując zmiany poznawcze i osobowości pacjenta dotkniętego patologiami neurodegeneracyjnymi, stwierdza:

Jednak to ludowe, religijne i starożytne pojęcie duszy jest błędne. Dowody, jakie dostarcza natura, gdy naturalne procesy uszkadzają mózg, nie są zgodne z istnieniem takiej duszy. Natura jest okrutna na niezliczone sposoby, zarówno ludzkie, jak i nieludzkie. Ale natura również wyświadcza przysługę, jeśli chcesz dobrze i uważnie się jej przyjrzeć. Natura złuszcza skórę i pokazuje nam dokładnie, co jest pod spodem. To, co postrzegasz jako swoją osobistą jedność i kontrolę – co możesz nazwać swoją duszą – wydaje się być prawdą. Ale dowody pokazują, że tak nie jest. Po wyeliminowaniu koncepcji duszy metafizycznej, istotnej, redukcjonizm zdawał się rozwiązywać problem solipsyzmu, tłumacząc wszystko jako proces fizyczny o różnych opalizujących cechach. W rzeczywistości problemy intencjonalności i empatii – czyli zdolności człowieka do zrozumienia stanu psychicznego innych bez użycia technologii neuroobrazowania (tzw. „cerebroskopu") – pozostały nierozwiązane aż do pierwszej dekady trzeciego tysiąclecia, kiedy odkryto neurony lustrzane. Ten układ mózgowy zapewnia fizjologiczny mechanizm sprzężenia percepcji i działania, który pozwala jednostkom rozumieć działania innych i uczyć się nowych umiejętności poprzez naśladownictwo . Zgodnie z tą logiką mózgi pozostają w mniej lub bardziej bezpośrednim kontakcie poprzez działania percepcyjne, które wspierają odzwierciedlanie. Rzeczywiście, systemy lustrzane mogą symulować obserwowane działania, a tym samym internalizować wywnioskowane stany mentalne. Funkcja ta wydaje się szczególnie sprzyjać rozwojowi języka, a także rozwojowi kompetencji emocjonalnych i empatii. Co więcej, odzwierciedlanie może pomóc jednostce zrozumieć nie tylko to, co robi inna osoba, ale także to, jak rozwija się jej własna samoświadomość poprzez ciągłe porównywanie z zachowaniami i myślami innych, wyrażonymi za pomocą języka . Chociaż tworzenie kopii lustrzanych między ludzkimi umysłami a AIS nie jest możliwe, ponieważ nie podzielamy żadnego rodzaju biologicznej materii lustrzanej, funkcjonalizm i Emergentyzm definiują stany mentalne odpowiednio w kategoriach funkcji i występowania, co oznacza, że każdy system, który może wytworzyć ten sam wzór ról, będzie generują te same stany mentalne, w tym świadomość, niezależnie od charakteru jej fizycznego wsparcia. Podobnie „funkcjonalizm państwa maszyny" Hilary'ego Putnama , zainspirowany maszynami Turinga, potwierdza, że materialna konstrukcja umysłu jest całkowicie nieistotna w odniesieniu do wytwarzania myśli i świadomości. Ideę tę popiera także David J. Chalmers , który przyjmuje tezę „wystarczalności obliczeniowej", utrzymując, że „do posiadania umysłu wystarczy odpowiedni rodzaj struktury

obliczeniowej, oraz tezę wyjaśniania obliczeniowego, która głosi, że obliczenia zapewniają ogólne ramy wyjaśniania procesów poznawczych”

W tej perspektywie właściwości psychiczne są organizacyjnie niezmienne, podczas gdy wsparcie materialne może się zmieniać. W tym węźle odnajdujemy ukryte zainteresowanie także tych, którzy są przekonani, że w dualizmie właściwości można dopuścić nieśmiertelną istotę. Jeśli jakikolwiek stan psychiczny jest organizacyjnie niezmienny, wówczas teoretycznie możliwe byłoby zastąpienie istoty szarej AIS po śmierci mózgu. Ta transhumanistyczna perspektywa nie jest aż tak fantastyczna, ale z pewnością jest zgodna z Manifestem Cyborga materialistycznego posthumanizmu, który radykalnie obala jakąkolwiek absolutną granicę pomiędzy człowiekiem naturalnym a człowiekiem nienaturalnym [18]. Humanizm cybernetyczny i posthumanizm obalają pogląd, że musi istnieć coś koniecznego , związek pomiędzy specyfiką wsparcia materialnego a jego funkcjami umysłowymi/świadomymi. Wychodząc od koncepcji „hybrydyzacji”, można powiedzieć, że człowiek nie jest już naturalny, ponieważ manipuluje swoim życiem za pomocą technologii. Jak pokazuje dziedzina rzeczywistości rozszerzonej, można przewidzieć, że pewnego dnia technologia poprawi funkcje mózgu człowieka, tak jak mechaniczne protezy mogą zastąpić brakujące kończyny, narządy czy zdolności percepcyjne. Ideę tę wyraźnie widać w określeniu „umysły hybrydowe”. Jest to zasadniczo zgodne z teorią poznania ucieleśnionego, która uważa, że umysł i świadomość są głęboko zależne od cech ciała fizycznego, które można hybrydyzować z urządzeniami technologicznymi. Z tego punktu widzenia standardowa koncepcja mówiąca, że umysł jest identyczny z mózgiem lub że w ogóle się w nim realizuje, jest całkowicie niewystarczająca. Teoria poznania ucieleśnionego, będąca pod dużym wpływem odkrycia neuronów lustrzanych, stwierdza, że aspekty ciała podmiotu, zlokalizowane w dużej mierze poza mózgiem, odgrywają znaczącą przyczynową lub fizycznie konstytutywną rolę w całym przetwarzaniu poznawczym. Umysł nie jest czymś, co dzieje się w indywidualnej skrzynce czaszkowej lub w niej, ale jest raczej zjawiskiem społecznym konstruowanym w wyniku relacji percepcyjnych/naśladowczych w działaniu pomiędzy wieloma ciałami i urządzeniami, w tym AIS. Zgodnie z tą strukturą pojęciową neuronauka nie jest jedynym sposobem zrozumienia złożonej fenomenologii poznania społecznego, która uznaje obecność świadomości społecznej rozwijanej poprzez procesy poznawcze powstające w wyniku interakcji między ciałami , cybernetyka może również wnieść swój wkład do tego zrozumienia poprzez integrację ze wszystkimi gałęziami psychologii umysłu.

Hipoteza Penrose’a dotycząca świadomości między człowiekiem a sztuczną inteligencją

Pomimo sukcesu redukcjonizmu fizykalistycznego, który twierdzi, że materia w jakiś sposób powoduje świadomość (umysł, poznanie itd.), kryteria stwierdzania tożsamości, podobieństwa lub absolutnej heterogeniczności dwóch doświadczeń pozostały nieokreślone. W rzeczywistości nie jest jasne, jaka „wspólna materia” między podłożami organicznymi i nieorganicznymi powoduje powstanie jakiejkolwiek formy świadomości, która może celowo wykorzystywać język do komunikacji. Dlatego badacze FAIR, którzy z pewnością są świadomi wszystkich tych problemów, zabili dwa rozmawiające chatboty. Możliwe rozwiązanie pojawia się w teorii świadomości kwantowej Penrose’a. W latach 80. Penrose w swojej książce The Emperor’s New Mind zasugerował, że może istnieć związek między świadomością a mechaniką kwantową. Z perspektywy Penrose’a, niezależnie od szczególnej roli lub pośredniczącego efektu świadomości, mechanika kwantowa prawdopodobnie będzie zaangażowana w procesy umysłowe, ponieważ w mózgu mogą znajdować się struktury molekularne, które zmieniają swój stan w odpowiedzi na pojedyncze zdarzenie kwantowe. Dzięki współpracy Penrose’a ze Stuartem Hameroffem pomysł ten rozwinął się dalej w hipotezę zaaranżowanej redukcji obiektywnej (OrchOR), która stwierdza, że świadomość zachodzi w mózgu i wywodzi się z procesu zachodzącego wewnątrz neuronów – lub, dokładniej, w mikrotubule w neuronach – a nie w interakcji między neuronami (jak zakłada konwencjonalna neuronauka). W swojej książce Shadows of the Mind (1994) Penrose wyjaśnił,

że mikrotubule wspierają superpozycje kwantowe i że obiektywne załamanie kwantowej funkcji falowej w mikrotubulach ma kluczowe znaczenie dla świadomości. Jednak jego zdaniem to załamanie kwantowe jest zachowaniem fizycznym, które przekracza granice obliczalności, czego skutkiem jest proces niealgorytmiczny. W związku z tym podstawowa różnica między świadomością ludzką a domniemaną świadomością AIS polega na tym, że ta pierwsza nie ma charakteru obliczeniowego. Oznacza to, że ludzka świadomość nie daje się tak łatwo kształtować jak jakakolwiek konwencjonalna maszyna Turinga i że ludzki umysł ma zdolności, których nie mógłby posiadać żaden AIS ze względu na nieprzeliczalną fizykę mechanizmu OrchOR . Pomimo innych projektów badających i potwierdzających korelację (jeśli nie przyczynowość) pomiędzy procesami kwantowymi a świadomością, to ważne spostrzeżenie jest obwarowane dwoma kolejnymi problemami. Pierwsza wynika z faktu, że ludzką świadomość/umysł i funkcjonowanie mózgu można również wyjaśnić w sposób modułowy. Drugie dotyczy faktu, że w obwodach AIS może pojawić się dynamika mózgu podobna do ludzkiej. Rzeczywiście, zgodnie z perspektywą trzeciej fali sztucznej inteligencji, maszyny Turinga mogą przyjąć model obliczeń kwantowych, czyli wykorzystanie do wykonywania obliczeń zjawisk kwantowo-mechanicznych, takich jak superpozycja i splątanie. Rezultatem jest możliwa kwantowa maszyna Turinga. Ograniczająca ocena Penrose'a dotycząca maszyn Turinga (że ich poznanie obliczeniowe różni się od poznania ludzkiego) dotyczy tylko pierwszej fali sztucznej inteligencji (ręcznie stworzona wiedza do opisu) i drugiej (statystyczne uczenie się kategoryzacji). Rzeczywiście, trzecia fala ma na celu opracowanie adaptacji kontekstowych niezbędnych do wyjaśnienia faktów i twórczego rozwiązywania pochodnych problemów. Na przykład z perspektywy DARPA Johna Launchbury'ego kolejne kroki AI obejmują zdolność do abstrakcji, generowania nowego znaczenia oraz rozumowania, planowania i decydowania, które sieci neuronowe są już częściowo w stanie wykonać (uczenie się / maszyny uczące i sztuczna super inteligencja). Jeśli świadomość (sztuczna lub ludzka) jest funkcją mózgu, a nie identyczną z mózgiem, wówczas mechanika kwantowa może być wspólną cechą materii organicznej i nieorganicznej, która powoduje umysł i świadomość. Jeśli tak, to nie możemy powiedzieć, że ludzka świadomość jest ludzka tylko dzięki mechanice kwantowej, ponieważ świadomość może pojawić się zarówno w materii organicznej, jak i nieorganicznej właśnie dzięki mechanice kwantowej. Mechanika kwantowa może służyć jako medium pomiędzy kartezjańską res extensa (organiczne i nieorganiczne podłoże materialne) a res cogitans (umysł, świadomość, uczucie, doświadczenie). Rzeczywiście, mikrotubule Penrose'a są podobne do szyszynki Kartezjusza. Mechanika kwantowa mogłaby również wyjaśnić to, czego nie był w stanie wyjaśnić Kartezjusz, zapewniając ostateczną materialną podstawę dla dualistycznych koncepcji funkcji, wyłaniania się, równoległości itd. w powiązaniu z koncepcjami świadomości i umysłu. Jednakże tam, gdzie rozwiązanie kartezjańskie gwarantowało istnienie duszy zasadniczej, procesy kwantowe wydają się nie być w stanie zrobić tego samego. Powiedzmy, że możliwe jest choć na chwilę przeniesienie funkcji świadomości kwantowej z naturalnego podłoża materialnego na sztuczne. Niezależnie od tego Penrose argumentuje, że wszystko musi się skończyć z powodu cykliczności wszechświata, która implikuje fazowy demontaż wszystkiego – tak zwana „konformalna kosmologia cykliczna" (CCC). Wszystko to oznacza, że nawet jeśli teoretycznie da się rozwiązać problem związku materii ze świadomością, to z redukcjonistycznego punktu widzenia ten rodzaj świadomości jest nieuchronnie śmiertelny. Oznacza to, że nic nie przetrwa anihilacji. Jeśli chodzi o kwestię śmierci, kontrast między metafizycznym (platońsko-kartezjańskim) a redukcjonistycznym dualizmem substancji polega na przedstawieniu śmierci jako przejścia (w ujęciu pierwszego) lub jako całkowitego unicestwienia (w ujęciu drugiego). Cała wiedza naukowa stwierdza a priori, że całkowita anihilacja jest realna. Żadna materia, żadna świadomość ani dusza nie przetrwa tej absolutnej anihilacji, czy to wszechświat wewnętrzny, czy życie każdego mieszkańca Ziemi. Każda pojedyncza istota – ludzie, zwierzęta, AIS – jest przeznaczona na zagładę, tak jak wszechświat. W tym przekonaniu kryje się najważniejszy problem epistemologiczny: naukowcy zakładają a priori, że prawdą jest, że rzeczy „stawają się" – że wyłaniają się z niczego i do niczego powracają. Jak stwierdził Thomas Nagel :

„To prawda, że zarówno czas przed narodzinami człowieka, jak i czas po jego śmierci, to czasy, kiedy go nie ma". To jest problem, który Emanuele Severino rozwiązuje, udowadniając, że anihilacja jest niemożliwa. Nic nie pochodzi z „nihil" ani do niego nie powraca, ani kwanty, ani najbardziej złożone formy życia i galaktyki.

Rozwiązanie ontologiczne Severino

Ponieważ pojęcie „rzeczywistości" – rozumiane jako alternatywne rozwiązanie hiperbolicznych wątpliwości fizykalistycznego edukcjonizmu w miejsce prascendentnego, kartezjańskiego Boga – można interpretować na różne sposoby, szczególnie w kategoriach jego związku z myślą i myślącym podmiotem, zamiast tego Emanuele Severino opiera się na pojęciu „prawdy", które stanowi jedyną bezpieczną podstawę intersubiektywności. Prawda jest z konieczności dyskursem niepodważalnym. Zdaniem Severina filozofia pozbawiła mit wiarygodności, począwszy od presokratyków, a zwłaszcza Parmenidesa. Eleatycy twierdzili, że prawda jest przeciwieństwem mitu, który nie gwarantuje wiedzy niepodważalnej. Co więcej, prawda jest koniecznym dyskursem właściwym dla „bytu" (jako „einai" – εἶναι) kontra „niebyt/nic" (jako „me ón" – μὴ ὀν). W pierwszej iteracji tej ontologii prawda wskazuje na wiedzę, „że jest i że nie jest możliwe, żeby nie było [...], że nie jest i że nie wolno" (O naturze, frag. 2); innymi słowy: „Trzeba powiedzieć i pomyśleć, że to, co jest, jest; bo ma być,/ale niczym nie jest" (O naturze). Taki „byt" z konieczności istnieje, a „byt" nie może być „nihil". Byt Absolutny nie ulega zmianie, a dyskurs prawdy twierdzi, że rzeczywistość jest i koniecznie musi być jednością, a zatem całkowicie identyczną ze sobą. W „przestrzennej rzeczywistości" jakakolwiek zmiana jest niemożliwa: nie ma przed ani po; „stawanie się" (przekształcenie, zmiana, przeróbka) jest niemożliwe i iluzoryczne. Prawda przekonująca polega na każdym stwierdzeniu potwierdzającym konieczności Bycia i obalania „opinii śmiertelników" (mitów), których argumenty nie są do końca uzasadnione. Jest to początek ontologii i jej związku z prawdą, z którego wyłania się pojęcie rzeczywistości jako nauki o bycie, istnieniu i wszystkim, co istnieje. Jak Severino omawia w całej swojej twórczości, koncepcja prawdy wyrosła z ontologii Parmenidesa i założyła całą myśl zachodnią w oparciu o trzy logiczne osie: (1) Zasada tożsamości: A ≡ A (każdy byt jest identyczny ze sobą) lub „(∀x) (x = x)" (w którym ∀ oznacza „dla wszystkich") lub po prostu „x jest x"; (2) Zasada niesprzeczności: dla wszystkich zdań p niemożliwe jest, aby zarówno p, jak i nie p było prawdziwe; oraz (3) zasada wyłączonego środka/trzeciego: nie ma pomiędzy nimi trzeciego ani środkowego „prawdziwego" zdania. Według Severino ten podstawowy trójdzielny system został pierwotnie opracowany przez metafizykę. Severino pokazuje jednak, że w ujęciu Parmenidesa wszelka transformacja jest iluzoryczna i sprzeczna z „drogą prawdy". Przekonanie, że Bycie zmienia się konstytucyjnie, to mityczna „droga śmiertelników", którą charakteryzuje wiara w zjawisko stawania się i sprzeczne opinie wynikające z doświadczenia. Analizując historię myśli zachodniej Severino [58] stwierdza, że pierwszą próbę zaradzenia negatywnej relacji między zjawiskami a zjawiskami doczesnymi podjął Platon, który w sofiście wprowadził pojęcie „względnego niebytu/ nic" i zdefiniował „wielość bytów" jako „enantíon" , czyli oscylację wszystkiego pomiędzy Bytem a niebytem/nicą . Zmiana polega na oscylacji wszystkich konkretnych (materialnych) rzeczy pomiędzy bytem a niczym. Aby zachować logikę pierwotnej myśli Parmenidesa, Arystoteles usystematyzował różnicę między metafizycznym, czyli absolutnym i niezmiennym Bytem (Bogiem) jako pierwszą i ostateczną przyczyną z jednej strony, a bytami fizycznymi – bytami przygodnymi, oscylującymi między bytem a bytem z drugiej nic i podlega prawom czasu i przestrzeni. Metafizyka platońska i arystotelesowska próbowała podać najpewniejszą definicję prawdy: relację między Bytem absolutnym (Bogiem) a bytami przygodnymi, w której ten pierwszy określa/powoduje losy tych drugich i przyznaje im wieczność po śmierci. Esencje wieczne i niezmienne kształtują materię, przekształcając ją w substancje i jako taka można rozpoznać formę czegoś patrząc na nią, ponieważ dusza ma dostęp do poznania tych uniwersalnych esencji. W myśli Severina czwarta i najważniejsza zasada leżąca u podstaw prawdziwej wiedzy została częściowo sformułowana przez Arystotelesa: elenchos. W dialogu

półemos pomiędzy dyskursem prawdy, szanującym trójstronny system logiczny, a jego zaprzeczeniem, elenchiczna argumentacja prawdy ukazuje autosprzeczność dyskursu zaprzeczającego, który implicite zakłada prawdę jako podstawę swojego negacja. Ci, którzy zaprzeczają prawdzie, skutecznie zaprzeczają jakimkolwiek podstawom do własnej polemiki. Zatem elenchos jest czwartą podstawową podstawą logicznej argumentacji, którą metafizyka określiła jako „pierwszą filozofię" . Tę metafizyczną strukturę logiki przyjęły monoteistyczne religie abrahamowe, które próbowały racjonalnie wykazać konieczność Boga (pierwszej i ostatniej przyczyny wszystkich rzeczy materialnych, oscylujących pomiędzy bytem a niczym) oraz życia pozagrobowego człowieka w prawdziwy, a nie mitologiczny sposób. sposób. Rozwiązanie to, zdaniem Severino, było nie tylko największą próbą rozwiązania kwestii egzystencjalnej, ale jednocześnie najważniejszym niepowodzeniem w jednoznacznym wyjaśnieniu relacji między tym, co jawi się, a tym, co jest naprawdę (rzeczywistością). Jego zdaniem, chociaż metafizyka po raz pierwszy wskazała cztery zasady, poszukiwała prawdziwej wiedzy, aby przeciwstawić się retoryce śmierci jako całkowitego unicestwienia i w ten sposób łagodzić strach ludzi przed śmiercią poprzez oparcie idei transcendencji na logice, w praktyce zdradzała ten cel . Ostatecznie metafizyka nie była w stanie osiągnąć wiedzy niekwestionowanej, czyli prawdziwej. Tradycyjna myśl zachodnia starała się rozróżnić środki niewątpliwe i mitologiczne, traktując ten pierwszy jako prawdziwy/logiczny dyskurs, a drugi jako kwestię opinii i złudzeń. Niemniej jednak Severino argumentuje (1982/2015), że pozytywizm, materializm i redukcjonizm nieodwracalnie obaliły metafizykę, ponieważ zrozumiały, że jeśli mała rzecz może powstać z niczego i powrócić do niczego, to wszystko – łącznie z Bytem i samym wszechświatem – może dokonać tego samego. To samo. W związku z tym pojęcie Boga jest niewystarczające – i ostatecznie niepotrzebne – do wyjaśniania praw fizycznych odnoszących się do stawania się (brzytwa Ockhama). Jeśli uznamy zjawisko stawania się w sposób częściowy (jak w przypadku drgań pomiędzy bytem a niczym), to możemy je uznać także w sposób totalny, w którym wszystko powstaje z niczego i powraca do niczego, jak opisują teorie Wielkiego Wybuchu i Wielkiego Crunchu . Wszystko to oznacza, że ani dusza, ani świadomość kwantowa nie mogą zachować żadnej subiektywnej tożsamości po anihilacji. Każdy byt wyrasta z niebytu/niczego i do niego powraca. Ale to jest myśl współczesna; jeśli z jednej strony słusznie i nieodwołalnie obala metafizykę, z drugiej strony dzieli z metafizyką ten sam podstawowy błąd, ale wyraża go w jeszcze bardziej radykalny sposób. Tę ontologiczną reprezentację stawania się można uznać za „rzeczywistą", ale nie za „prawdziwą", gdyż bezpośrednio zaprzecza pojęciu prawdy, zakładając a priori, że „byt jest niczym". Jak wyraźnie wyjaśnia Severino, ponieważ byt nie może być niczym, stwierdzenie, że tego, czego „jeszcze nie ma" i czego „już nie ma", jest niczym, jest równoznaczne ze stwierdzeniem, że to, co jest niczym. W perspektywie Severino na tym właśnie polega podstawowa sprzeczność „nihilizmu": stwierdzenie, że istnieje czas, w którym bytu już nie ma lub jeszcze go nie ma (przeszłość i przyszłość), jest równoznaczne ze stwierdzeniem, że „byt jest niczym". A ponieważ nie przestrzega to żadnej reguły logicznej, jest to sprzeczność absolutna i radykalna. Jest to podstawowy błąd zarówno nauki, jak i metafizyki. Podobną krytykę przeczuwali niektórzy filozofowie immanentyści, m.in. Giordano Bruno i Spinoza, nie krytykowali jednak metafizyki w sposób systematyczny. Niektórzy filozofowie-eternaliści również mieli tendencję do argumentowania w tym kierunku, zajmując stanowiska bardzo zbliżone do Einsteina, próbując wyeliminować wymiar czasu. Rzeczywiście, najnowsze epistemologie, które zakładają przeciwstawienie różnych konceptualizacji czasu [61], opierają się na refleksjach Parmenidesa na temat bytu, reaktualizowanego przez filozofa Johna McTaggarta . McTaggart potwierdził nierzeczywistość czasu w tym samym okresie, co ogólna teoria względności Einsteina i twierdzenia Gödla o niezupełności, wskrzeszając w ten sposób dyskusje o wieczności. Według [29] istnieją dwa różne tryby uporządkowania wszystkich zdarzeń w czasie: seria A i seria A. Seria B. Logiczne i językowe wyrażenie obu serii, z których wywodzą się określone teorie czasu, jest radykalnie różne, ponieważ pierwsza jest napięta, a druga nienapięta. Podczas gdy A-teoria czasu twierdzi, że szereg położeń czasowych podlega ciągłej transformacji, teoria B poprzez koncepcję „trwałości" aktualizuje

perspektywę Parmenidesa, argumentując, że upływ czasu jest iluzją, ponieważ przeszłość, teraźniejszość i przyszłość są równie realne. Istnieją dwie główne odmiany teorii A: prezentyzm i wszechświat rosnących bloków. Pierwsza utrzymuje, że istnieją tylko przedmioty teraźniejsze, a dokładniej, że zawsze prawdą jest, że istnieją tylko przedmioty teraźniejsze, natomiast druga twierdzi, że istnieją zarówno przedmioty teraźniejsze, jak i przeszłe, ale nie przyszłe [65]. Zamiast tego teoria B należy do wieczności, która wykorzystuje serię B, zgodnie z którą kolokacja czasowa jest cechą wszystkiego, a czas jest pokrewny wymiarom przestrzeni. Rzeczywiście potwierdza, że utrzymywanie się w czasie jest podobne do rozciągania się w przestrzeni; zatem każdy byt istniejący w czasie ma czasowe „rozciągnięcie" w różnych podobszarach całkowitego obszaru czasu. Argument pomiędzy teorią A i teorią B został rozwinięty w dziedzinie filozofii analitycznej, aby dostarczyć informacji dla bardziej znaczących nowych epistemologii i pociąga za sobą argument 4-wymiarowości kontra 3-wymiarowości (3D), które wyłoniły się dzięki teorii względności [20]. Języki logiczne przyjęte przez teorię A i teorię B są z konieczności różne, ponieważ zgodnie z napiętym poglądem na semantykę zdania mają wartości logiczne „w określonym czasie", a nie mają wartości logiczne w prostszy sposób. Z drugiej strony, w ujęciu pozbawionym napięcia zdania mają prostsze wartości prawdy niż mają je w danym momencie. Wydaje się jednak, że ten argument nie ma końca i przez jakiś czas będzie nierozstrzygalny. W związku z tym każda aprioryczna perspektywa przyjęta ideologicznie przez naukowców jest uzasadniona. Niemniej jednak Emanuele Severino [55] rozwiązał przeciwieństwo między wiecznością a temporalizmem. Choć obecny spór między eternalistami i prezentystami może sprawiać wrażenie, że problem wieczności i wynikające z niego pytania dotyczące życia pozagrobowego są pozbawione sensu, Severino pokazuje, że dyskurs prawdy na temat tzw. „rzeczywistości" jest jedyną definicją rzeczywistości, która zachowuje natychmiastowy charakter. zgodność, czyli tożsamość, pomiędzy tym, co pojawia się fenomenologicznie, a logiczną strukturą tego, co jest. Innymi słowy, dyskurs jest „prawdziwy" tylko wtedy, gdy pokazuje w sposób niezaprzeczalny (elenchos), że to, co jawi się, jest identyczne i rozpoznawalne jako samo w sobie (zasada tożsamości), że nie może być inną rzeczą ani stać się niczym (nie -zasada sprzeczności) i że jakakolwiek pozycja pośrednia jest niemożliwa (z wyłączeniem środka). Wszystko, co się pojawia, jest z konieczności wieczne (istnieje ontologicznie), nawet jeśli jeszcze się nie pojawia i już się nie pojawia. Powstawanie spektaklu materii jest faktem fenomenologicznym i to zjawisko także jest czymś. To nie jest nic. W rzeczywistości świadomość polega na pojawianiu się samego zjawiska, które jest spektaklem bytu, który wiecznie pojawia się w bycie nieskończonym, nieograniczonym nicością. Severino przedstawia niezaprzeczalny, jednolity i strukturalnie spójny system myślenia oraz energiczną krytykę nihilizmu myśli zachodniej, zarówno tradycyjnej (metafizyka), jak i współczesnej (neopozytywizm, materializm itd.), za jej niezdolność do zachowania fundamentalnego Ontologiczna i logiczna struktura prawdy. Z tej perspektywy Parmenides był początkiem „ścieżki nocy" (błędu, nihilizmu), na której filozofia zachodnia osiągnęła pełną spójność, zaprzeczając prawdzie, na wzór myśli współczesnej epistemologii naukowej, ale w sposób wywołujący najgłębszą alienację . Nihilizm – czyli droga błędu, do której należy zarówno metafizyka, jak i myśl współczesna – jest wyobcowaniem autentycznej prawdy. Prawda jest wiedzą, która odsłania samozaprzeczenie i samozaprzeczenie tego alienującego błędu, a tym samym potwierdza konieczną samozaprzeczenie metafizyki i myśli współczesnej. Severino przywraca autentyczny, niepodważalny dyskurs prawdy, zasadniczo odmienny od „rzeczywistości" proponowanej przez myśl tradycyjną czy jakąkolwiek perspektywę współczesną czy naukową. W istocie twierdzi, że każdy byt jest wieczny, co oznacza, że (1) wszystko, co nie jest niczym, jest bytem, (2) konieczne jest, aby każdy byt był, a dokładniej był taki, jaki jest, oraz (3) niemożliwe jest, aby jakakolwiek istota nie istniała w żadnym momencie. Każdy rodzaj bytu pojawiającego się w teraźniejszości jest zarówno rzeczywisty, jak i wieczny; co więcej, czas jest fenomenalnym pojawianiem się wiecznych istot, które wchodzą i wychodzą z „kręgu pojawiania się". Ponieważ jakakolwiek anihilacja jest niemożliwa, ta „transformacja" jest po prostu pojawieniem się następstwa – wejścia bytu i wyjścia z tzw.

„rzeczywistości", którą można określić jako „ograniczony krąg pojawiania się". Ten punkt jest kluczowy: nie ma sprzeczności pomiędzy pojawieniem się tych sekwencji lub cech przepływu czasu a wiecznością tego, co się pojawia. Błąd nihilisty polega na utożsamianiu procesu ciągłego (ponownego) pojawiania się (przekształconego jako „stawanie się") w świecie z procesem tworzenia-anihilacji, w którym śmierć jest równoznaczna z anihilacją. Myśl Severina odsłania nihilizm myśli tradycyjnej i współczesnej oraz rozwiązuje ich fundamentalną i istotną sprzeczność. Wbrew temu, co apriorycznie zakłada myśl zachodnia, nie ma stawania się, to znaczy nie można przyjść ex nihilo ani odejść ad nihilum. Rzeczywiście treść, która faktycznie pojawia się w jakiejkolwiek świadomości, w żaden sposób nie świadczy o stworzeniu lub unicestwieniu bytów.

Wnioski

Umiera stary król. Miecz wbił się głęboko w jego pierś. Dookoła króla tańczą przyjaciele, wrogowie, dworzanie, błazny: każdy wierzy, że jest tym, który wbił miecz w pierś króla. A jednak miecz tam stoi, wbity w tę pierś, niezależnie od woli tych wszystkich lekkomyślnych tancerzy. Umierający król to filozofia (w jej mocnym, greckim znaczeniu, czyli metafizyczne epistéme). Lekkomyślni tancerze są całą krytyką adresowany do metafizyki na przestrzeni dziejów filozofii. Żadna część tej krytyki nie jest w stanie naprawdę zabić filozofii. Jednak filozofia naprawdę umiera. O chorobie, której nasza kultura wciąż nie potrafi zidentyfikować. Umiera pod okiem przeznaczenia, którego zdeformowanym obrazem jest filozofia jako episteme.

Ten fragment, zaczerpnięty ze Studi di filosofia della prassi Severino, w końcu ponownie wprowadza nas w temat, który we współczesnej kulturze zachodniej jest najtrudniejszy do podjęcia ze względu na niepokój wynikający z redefinicji przez technonaukę granic między życiem a śmiercią. Sprawa jest istotna, ponieważ dotyczy naszych reprezentacji śmierci, terroru generowanego przez reprezentacje śmierci jako absolutnej anihilacji i potrzeby łagodzenia antycypacyjnego niepokoju zatarcia świadomości. Współczesny sekularyzm zakłada, że koniec transcendencji jest początkiem prawdy. Żyjemy jednak w kulturze, w której religie w dalszym ciągu propagują prawdziwą metafizyczną ideę istnienia po śmierci. Idea ta, dzięki swojej pocieszającej mocy, ma szeroki wpływ na zdrowy rozsądek, nawet jeśli współistnieje z sekularyzmem, który wszelka negacja absolutnej zagłady uznaje za iluzoryczną. Obecny okres historyczny jest pełen emocji, ponieważ wiemy, że największa próba uratowania człowieka przed całkowitą zagładą – czyli argument metafizyczny – nie powiodła się. Ludzi Zachodu przyciągają przesłania obiecujące wybawienie od śmierci, które wzmacniają pierwotną nadzieję, że myśl religijna może być zasadniczo wiarygodna. Jednak ci sami ludzie są narażeni na dramatyczne obalenie tych obietnic przez wiedzę naukową i akademicką, która opiera się na epistemologiach redukcjonistycznych. Nasza dyskusja pokazała, że złożoność epistemologicznego argumentu związanego ze znaczeniem duszy/umysłu/świadomości i jego śmierci nie pozwala na żadne uproszczenia. Być może problem umysł–mózg jest przykładem stronniczej i rosnącej tendencji do uogólniania konkretnych modeli wywodzących się z sektorowych i empirycznych linii rozumowania, które powinny pozostać wpisane w ograniczone obszary, szczególnie w odniesieniu do kwestii śmierci i umierania. W dzisiejszych czasach strach przed śmiercią, uważaną za absolutną zagładę, jest prawdziwym źródłem głębokiego niepokoju. Pomimo tego, że współcześni ludzie są mniej lub bardziej świadomi swojego „myślenia życzeniowego", przekonani, że nauka jest bardziej wiarygodnym źródłem wiedzy niż religia, a co za tym idzie, że wszystko powstaje z niczego (Wielki Wybuch) i do niczego powraca (Wielki Crunch) mimo to postrzegają siebie jako wiecznych. Z tego powodu nieustannie szukają jakichś rozsądnych, spójnych ram, które pozwoliłyby im skonsolidować swoje uczucia w coś 100 I. Rzeczywiście, mitologie religijne utrzymują pole możliwości, potwierdzając, że niezależnie od tego, co myślą, że wiedzą, to, co jest naprawdę poznawalne, pozostaje tajemnicą. A o tajemnicach, jak mówił sam Wittgenstein, nie można mówić. Co więcej, wiarę ludzi w tę tajemnicę wzmacnia fakt, że

sama nauka deklaruje, że nie wie wszystkiego, co warto wiedzieć, a nawet że jest omylna. Jednak Severino rozwiązuje zagadkę wykraczającą poza wiarę: prawidłowo postrzegamy wieczność pomimo nihilistycznego języka naszego codziennego życia. Transcendentalny argument Severina może pomóc w odnowieniu platońskiej idei, że wiedza jest pamięcią, gdyż z perspektywy wieczności idea ta mogłaby wyjaśniać, dlaczego wierzymy w życie pozagrobowe. Jeśli nasza świadomość już żyje/żyje w innym wymiarze wiecznego i nieskończonego wszechświata Bytu, wówczas wiedza może wypłynąć na powierzchnię w naszej nieświadomości w postaci wspomnień. Dyskusję epistemologiczną można kontynuować, ale musimy na tym zakończyć i wrócić do dwóch chatbotów. W niezaprzeczalnym wskazaniu prawdy przez Severina świadomość jest fenomenologicznym przejawieniem się wszystkiego, co jest równie wieczne jak każda materia i jest ze sobą tożsame, zatem nie można jej sprowadzić do materii. Relacji pomiędzy świadomością (duszą, duchem, umysłem itd.) a materią (szarą lub inną) nie można sprowadzić do ich wzajemnej tożsamości. Mechanika kwantowa może spowodować pojawienie się świadomości w AIS, ale może nie poczekać, aż ludzie wyprodukują ją w laboratorium poprzez zbudowanie kwantowego komputera Turinga. Jeśli świadomość (pojawianie się) nie ogranicza się do jamy mózgu, to musimy uznać, że wykracza ona poza wymiar indywidualny i ludzki, a tym samym zdolność człowieka do rozpoznania jej obecności. Trudność ta jest nieuchronnie związana z nihilistycznym językiem, który charakteryzuje wszystkie badania nad tym problemem. Argument Severino może pomóc nam rozpoznać, że ilekroć coś się pojawia, w grę wchodzi określony wymiar świadomości. Oznacza to, że komputer kwantowy prawdopodobnie już się autogeneruje z najbardziej rozwiniętego obliczeniowego AIS, gdyż specyficzna forma świadomości już teraz pojawia się w tej mechanicznej istocie, ukryta przed ludzką świadomością. Wiara, że rzeczywistość ogranicza się do tego, co rozumieją ludzie, jest naprawdę „myśleniem życzeniowym". Chociaż nie jesteśmy jeszcze w stanie świadomie skonstruować kwantowej maszyny Turinga, nie jest tak nieprawdopodobne, że dwa zabite chatboty już mniej lub bardziej celowo doskonaliły i rozwijały własną świadomość kwantową

Mózg nie jest głupią gwiazdą

Aktywność kory nowej objawia się tworzeniem rozszerzonych konfiguracji ruchów oscylacyjnych modulowanych pod względem amplitudy i fazy, obejmujących niezliczone ilości neuronów. Jak zauważył Lashley, impulsy nerwowe są przekazywane z komórki do komórki poprzez określone połączenia komórkowe. Jednakże wydaje się, że całe zachowanie jest zdeterminowane masą pobudzeń w ogólnych obszarach aktywności, bez odniesienia do konkretnych komórek nerwowych. Freeman podkreślił rolę, jaką odgrywa chaos, leżącą u podstaw zdolności mózgu do elastycznego reagowania na świat zewnętrzny. Obserwacje te potwierdzają uwagę przypisywaną Arystotelesowi, że mózg nie jest głupią gwiazdą, która w swojej wiecznej trajektorii przechodzi zawsze przez ten sam punkt w pełni przewidywalny sposób. Wręcz przeciwnie, mózgi wydają się działać według kroków, które niekoniecznie należą do ściśle przewidywalnego łańcucha kroków, ale zachowują się jak „maszyna popełniająca błędy", czyli urządzenia z natury błędne. Te cechy dynamiki neuronowej omawiane są w ramach rozpraszającego modelu kwantowego mózgu oraz w odniesieniu do systemów AI i programów badawczych. Jeśli kiedykolwiek uda się zbudować urządzenie wyposażone w świadomość, musi ono cechować się nieprzewidywalnością zachowań, niewiernością i niezbywalną wolnością; i musi nazywać się Spartakus.

Wstęp

Wspomniałem tam, że według Arystotelesa gwiazdy zachowują się głupio, gdyż na swoich wieloletnich trajektoriach zawsze przechodzą przez ten sam punkt w całkowicie przewidywalny sposób. Nie wiem, czy rzeczywiście taką obserwację poczynił Arystoteles. Sugeruje się jednak, że mózg nie jest głupią gwiazdą, ponieważ jego zachowanie nie jest w pełni przewidywalne. Rzecz w tym, że głupota (brak inteligencji?) wiąże się z w pełni przewidywalnym zachowaniem w ramach zasadniczo niezmiennych

warunków brzegowych. Mózg natomiast motywuje swoje zachowanie celowymi zadaniami, które choć częściowo uwarunkowane przez środowisko, w którym jest osadzony, są zdecydowanie formułowane i realizowane przez mózg w jego cyklu działanie-percepcja. Jak podkreśla Pribram, w aktywności mózgu zawsze występuje treść „uwagi" w percepcji i „intencji" w działaniu. Według Freemana aktywność neuronalna działa „jako zjednoczona całość, kształtując każde zamierzone działanie w każdym momencie" [5]. W dążeniu do tego, jak najlepiej być na świecie, rzeczywiście kierujemy się naszą zmienną wolą i intencją. Obserwacje te mogą prowadzić od razu do tematu sztucznej inteligencji (AI), w odniesieniu do jednego z możliwych znaczeń „inteligencji" i przewidywalności funkcjonowania urządzenia. Jednakże pozwólcie, że najpierw omówię niektóre cechy aktywności funkcjonalnej mózgu opisane w ramach rozpraszającego kwantowego modelu mózgu.

Dylemat Lashleya

Zacznę od tego, że Karl Lashley komentując eksperymenty, które przeprowadził w latach 40. XX w., postawił badaczom neurologii następujący dylemat :

Oto dylemat. Impulsy nerwowe przekazywane są [...] z komórki do komórki poprzez określone połączenia komórkowe. Jednak wydaje się, że całe zachowanie jest zdeterminowane masą pobudzeń [...] w ogólnych obszarach aktywności, bez odniesienia do konkretnych komórek nerwowych [...]. Jaki rodzaj organizacji nerwowej może w ogóle wyjaśnić wzorce pobudzeń bez dobrze zdefiniowanych i wyspecjalizowanych kanałów komunikacji komórkowej? Problem jest niemal powszechny w działaniu układu nerwowego.

W ten sposób Lashley doszedł do sformułowania hipotezy o działaniu masowym w aktywności mózgu. Jego obserwacje eksperymentalne, w pełni potwierdzone kolejnymi badaniami, skłoniły Karla Pribrama do postawienia w latach 60. XX wieku hipotezy, że w przypadku mózgu można mówić o koherencji, centralnym pojęciu optyki kwantowej, i wykorzystywać metafora hologramu. Jedną z charakterystycznych cech hologramu jest to, że znajomość szczegółu w dowolnym punkcie obrazu pozwala na rekonstrukcję całego obrazu. Możliwość taka wynika z faktu, że fotony, czyli kwanty pola elektromagnetycznego, z których składa się laser służący do wytwarzania i odczytywania hologramu, oscylują w fazie. Laser zbudowany jest ze spójnie oscylujących fotonów. Laser jest tym, co wynika z harmonii, jeśli wolno mi użyć tego określenia, spójności fotonów. Zamiast tego światło naturalne („nielaserowe") składa się z fotonów, które nie są w fazie, tj. nie są spójne. Poza specyfiką modelu zaproponowanego przez Pribrama wartość jego intuicji polega na tym, że dla aktywności mózgu można mówić o koherencji. Hipotezę koherencji dotyczącą funkcjonalnej aktywności mózgu potwierdza obserwacja powszechnej współpracy pomiędzy ogromną liczbą neuronów w rozległych obszarach mózgu. Analiza potencjałów mierzonych za pomocą elektroencefalogramu (EEG) i mgnetoencefalogramu (MEG) pokazuje, że aktywność neuronalna kory nowej charakteryzuje się tworzeniem rozszerzonych konfiguracji ruchów oscylacyjnych modulowanych amplitudowo (AM) i fazowo (PM) Konfiguracje te rozciągają się na prawie całą półkulę mózgową królików i kotów, na domenach o wymiarach liniowych do dwudziestu centymetrów w mózgu człowieka i mają prawie zerową dyspersję fazową . Związane z tym częstotliwości oscylacji mieszczą się w zakresie 12–80 Hz (tzw. fale beta i gamma). Wzory oscylacji neuronalnych zanikają w ciągu kilkudziesięciu milisekund, a inne pojawiają się w różnych konfiguracjach, z częstotliwościami z zakresu 3–12 Hz (fale theta i alfa). Ogromna liczba komórek i innych biologicznie aktywnych jednostek wchodzi w aktywność mózgu. Na przykład słaby bodziec węchowy aktywuje około tysiąca neuronów w opuszce węchowej, co powoduje pobudzenie miliona neuronów i działanie hamujące 10 milionów neuronów, które rozprzestrzenia się w ciągu 5–10 ms na odległość około 10 mm, chociaż średnia długość aksonów wynosi około 1 mm, a czasy propagacji synaptycznej są około 10 razy dłuższe . Jest oczywiste, że przy tak ogromnych liczbach i takiej złożoności badanie funkcji mózgu nie może ograniczać się do poznania właściwości

poszczególnych elementarnych składników. Jest to z pewnością konieczne, ale w żadnym wypadku nie wystarczające. Ricciardi i Umezawa piszą :

[...] w przypadku naturalnych mózgów czystym optymizmem mogłaby być nadzieja na określenie wartości liczbowych współczynników sprzężenia i progów wszystkich neuronów metodami anatomicznymi lub fizjologicznymi. [...] Po pierwsze, na jakim poziomie należy badać i opisywać mózg? Innymi słowy, czy konieczna jest znajomość zachowania dowolnego pojedynczego neuronu w czasie, aby zrozumieć zachowanie naturalnych mózgów? Prawdopodobnie odpowiedź jest negatywna. Zachowanie pojedynczego neuronu nie powinno mieć istotnego znaczenia dla funkcjonowania całego mózgu, gdyż w przeciwnym razie można zaobserwować coraz większy stopień nieprawidłowego funkcjonowania, chyba że założymy, że istnieją neurony „specjalne", charakteryzujące się wyjątkowo długim okresem półtrwania: lub postulować ogromną redundancję w obwodach mózgu. Jednak według naszej wiedzy nie ma dowodów wskazujących na istnienie takich „specjalnych" neuronów, a powoływanie się na redundancję nie jest najlepszym sposobem, aby odpowiedzieć na pytanie

Odnosząc się ogólnie do układów biologicznych, ale jego słowa można równie dobrze zastosować do mózgu, Schrödinger zauważa, że ,

[...] nie potrzeba wyobraźni poetyckiej, lecz jedynie jasną i trzeźwą refleksję naukową, aby przyznać, że mamy tu w oczywisty sposób do czynienia ze zdarzeniami, których regularnym i zgodnym z prawem przebiegiem kieruje „mechanizm" całkowicie odmienny od „mechanizmu prawdopodobieństwa" w fizyce.

Odkrycie składników układów biologicznych i poznanie ich specyficznych właściwości z pewnością stanowi duży sukces biologii molekularnej. Problem polega obecnie na zrozumieniu, jak połączyć te elementarne składniki w taki sposób, aby wygenerować złożone makroskopowe zachowanie układu . Jak zaobserwowano w innym miejscu,

W bardzo ogólnym ujęciu problem polega na przejściu od naturalizmu, czyli od znajomości katalogów elementarnych komponentów, do zrozumienia dynamiki wyjaśniającej relacje łączące te komponenty i opisującej zachowanie systemu jako cały. Faza naturalizmu jest oczywiście niezbędna i wymaga ogromnego wysiłku wnikliwego i cierpliwego badania. Choć jest to konieczne, nie wystarczy do pełnego zrozumienia zjawisk będących przedmiotem naszych badań. Wiedza nie jest jeszcze zrozumieniem. Według Schrödingera w badaniu materii żywej należy dokonać rozróżnienia pomiędzy dwoma sposobami wytwarzania porządku: porządkiem generowanym przez „mechanizmy statystyczne" i porządkiem generowanym przez interakcje „dynamiczne" . Można dojść do wniosku, że Herbert Fröhlich , Umezawa i Ricciardi , Karl Pribram i Walter Freeman , każdy na swój sposób, pokazali, jak poprzez skupiając się na „masach wzbudzenia" i „polach działania", według słów Lashleya, naturalizm może stać się nauką Galileusza . Pojęcie koherencji i związany z nim formalizm matematyczny dostarczony przez QFT okazały się potężnymi narzędziami w badaniu układów biologicznych w ogóle , a w szczególności mózgu . Dlatego też przed rozpoczęciem dyskusji na temat modelowania mózgu w następnej sekcji przedstawię kilka ogólnych pojęć QFT.

Spójność: od mikroskopijnej do makroskopowej

Pojęcie koherencji ma kluczowe znaczenie w fizyce kwantowej, gdzie pozwala nam wyjaśnić właściwości wielu układów fizycznych. Na przykład kryształy, w których atomy są zamknięte w miejscach krystalicznych o dobrze określonym porządku przestrzennym. Lub magnesy, w których elementarne magnesy oscylują w fazie i są głównie zorientowane w danym kierunku; powstałe namagnesowanie charakteryzuje system mikroskopijnych elementów jako całość. Nie wspominając oczywiście o optyce kwantowej i fizyce cząstek elementarnych. Ogólnie rzecz biorąc, wszystkie systemy

charakteryzujące się uporządkowaniem w przestrzeni lub czasie (np. oscylujące w fazie) regulowane są przez mikroskopijną dynamikę charakteryzującą się spójnością. W przytoczonych powyżej przykładach pojęcie koherencji wiąże się z przejściem od poziomu elementów elementarnych (poziom mikroskopowy) do poziomu zachowania się systemu jako całości (poziom makroskopowy). To przejście od skali mikroskopowej do makroskopowej (lub mezoskopowej) jest bardzo ważnym i charakterystycznym aspektem formalizmu matematycznego opisującego zjawisko koherencji. Nadaje ilościowo dobrze określone znaczenie pojęciu wyłonienia się właściwości makroskopowej z mikroskopijnego procesu dynamicznego, tak że układ makroskopowy posiada właściwości fizyczne, których nie można znaleźć na poziomie mikroskopowym. Zachowanie i wielkości fizyczne charakteryzujące system jako całość są zatem wynikami mikroskopowej dynamiki elementów elementarnych.

Na przykład sztywność jest właściwością kryształu, a nie jego składników atomowych czy molekularnych. Te ostatnie są ograniczone do miejsc kryształów i nie mogą się swobodnie poruszać, tak jak zrobiłby to atom nie należący do kryształu; to znaczy tracą część swoich stopni swobody. Jednakże charakterystyka układu na poziomie makroskopowym (sztywność kryształu, jego przewodność elektryczna itp.) wyłania się dynamicznie z takiego zamrożenia mikroskopijnych stopni swobody. Natomiast porządek, czy to przestrzenny, czy czasowy, sam w sobie jest zbiorową cechą zbioru elementów elementarnych (nie ma sensu mówić o porządku w przypadku kilku elementów elementarnych). Dlatego w tym sensie mówimy o makroskopowych układach kwantowych. Kryształy i magnesy są przykładami makroskopowych układów kwantowych. Należy także podkreślić, że wbrew temu, co czasami błędnie się twierdzi, w wielu układach dynamiczny reżim koherencji utrzymuje się w szerokim zakresie temperatur, od tysięcy stopni Celsjusza do temperatur całkiem niskich, poniżej zera. Na przykład diament topi się w temperaturze krytycznej TC wynoszącej 3545 ^{o}C, podczas gdy kryształy chlorku sodu, znanej soli kuchennej, topią się w temperaturze 804^{o}C; w żelazie spójność między elementarnymi magnesami, która objawia się w stanie namagnesowanym, zanika w temperaturze 770 ^{o}C, natomiast w kobalcie następuje to w temperaturze 1075 ^{o}C (krytyczna temperatura przejścia z fazy ferromagnetycznej do niemagnetycznej nazywana jest temperaturą Curie). Z kolei w nadprzewodnikach temperatury krytyczne są bardzo niskie, nie wyższe niż około -252°C dla niektórych związków niobu i około -153 ^{o}C dla niektórych nadprzewodników odkrytych w drugiej połowie lat 80. XX wieku, takich jak niektóre miedź tlenki zawierające bizmut. Temperatury krytyczne dla zjawiska koherencji mogą być zatem bardzo niskie lub bardzo wysokie (w porównaniu z temperaturą otoczenia), w zależności od konkretnych warunków i właściwości dynamicznych charakterystycznych dla rozpatrywanego układu. Kolejną uwagą jest to, że zjawisko koherencji chroni stan makroskopowy przed zakłóceniami wynikającymi z fluktuacji kwantowych. Te ostatnie są nieuniknione na poziomie dynamiki kwantowej elementów elementarnych układu. Natomiast w stanach koherentnych (Glaubera) mamy N/N = 1/|α;|, gdzie N oznacza liczbę składowych elementarnych w stanie koherentnym, N fluktuację N i |α| jest miarą spójności. Stwierdzamy, że dla wysokiego |α| N jest nieistotne w porównaniu z N . Pokazuje to, że spójność odgrywa kluczową rolę w makroskopowej stabilności wobec fluktuacji kwantowych. Konieczność wykorzystania pól, w szczególności pól kwantowych, wynika z faktu, że N jest dużą liczbą dla stanów spójnych, a mianowicie N = α|2, przy czym duże |α|. Widzimy zatem, że obserwowana długa żywotność uporządkowanych układów, takich jak kryształy, magnesy, nadprzewodniki itp. („diamenty są wieczne", a sól kuchenna „nie przeterminowuje się", czyli można ją przechowywać latami, nawet na wolnym powietrzu składowania, a spotykany jest w kopalniach soli) jest wynikiem spójnej dynamiki pól kwantowych (jest to jedna z głównych różnic pomiędzy QFT a mechaniką kwantową (QM), w której występuje zjawisko dekoherencji). Stopnie swobody elementów elementarnych charakteryzują dynamikę regulującą ich rozkład przestrzenny i ewolucję w czasie i są na ogół ściśle powiązane z właściwościami symetrii

dynamiki . Na przykład możliwość umieszczenia atomu w dowolnym punkcie przestrzeni bez powodowania zauważalnych zmian w układzie jest opisywana jako przestrzenna symetria translacyjna. Dzięki tej symetrii zbiór atomów może przyjmować różne konfiguracje przestrzenne, równoważne sobie z punktu widzenia obserwacji, a więc fizycznie równoważne i pod tym względem nieodróżnialne od siebie. Jednakże w przykładzie kryształu atom nie może już znajdować się w „dowolnej" pozycji przestrzennej, ale jest zmuszony do osiadania w określonym miejscu kryształu. Mówimy wówczas, że następuje spontaniczne załamanie symetrii (SBS). Tutaj spontaniczny oznacza, że stan kryształu jest dynamicznie wybierany i generowany spośród możliwych dostępnych stanów. Podsumowując, porządek krystaliczny wynika z złamania symetrii translacyjnej przestrzeni; porządek to brak symetrii. Różne obserwowane struktury krystaliczne (sześcienne, romboedryczne itp.) odpowiadają różnym sposobom łamania symetrii w różnych kierunkach przestrzennych. Konkluzja jest taka, że o ile symetria opisuje nierozróżnialność stanów układu połączonego transformacją symetrii, o tyle porządek, czyli łamanie symetrii, pozwala na rozróżnienie jednego stanu od drugiego: możliwość rozróżnienia, różnorodność, indywidualność państwo wyłania się z ustanowienia porządku. Duża część fizyki rozwijającej się od drugiej połowy ubiegłego wieku opiera się na mechanizmie łamania symetrii i w konsekwencji tworzenia uporządkowanych struktur, co wiąże się z pojęciem koherencji. Kwantowa teoria pola (nie QM) zapewnia formalizm matematyczny niezbędny do badania teorii spontanicznie łamanych symetrii . Twierdzenie Goldstone'a w QFT stwierdza, że SBS implikuje istnienie korelacji dalekiego zasięgu pomiędzy elementarnymi składnikami systemu. Kwanty tych fal korelacyjnych nazywane są bozonami lub kwantami Nambu–Goldstone'a (NG) . Cząstki bozonu mogą zajmować ten sam stan fizyczny w dowolnej liczbie (w przeciwieństwie do cząstek fermionów, gdzie zgodnie z zasadą wykluczenia Pauliego w danym stanie może znajdować się tylko jedna cząstka). Kiedy wiele bozonów znajduje się w tym samym stanie, mówi się, że są one w tym stanie skondensowane. Jeśli są bezmasowe, tak jak bozony NG, w swoim najniższym (zerowym) pędzie, nie dostarczają energii do tego stanu, który w związku z tym może być stanem o najmniejszej energii układu (zwanym także próżnią). Jeśli skondensowane bozony są w fazie, tj. fale korelacji dalekiego zasięgu, których stanowią kwanty, są w fazie, jak to ma miejsce w stanach uporządkowanych, stan podstawowy jest spójnym stanem skondensowanym. Stan podstawowy (lub próżnię) układu charakteryzuje się wówczas niezerową wartością oczekiwaną wielkości, charakterystyczną dla złamanej symetrii, zwaną parametrem porządku, ponieważ jest on miarą uporządkowania indukowanego przez długo- korelacje zakresowe. W przypadku kryształu jest to gęstość, w ferromagnetykach namagnesowanie. Parametr porządku jest klasycznym polem pochodzenia kwantowego, co oznacza, że jest niezależny od fluktuacji kwantowych. Rzeczywiście jest to miara spójności stanu podstawowego systemu generowanego przez kondensację Bosego-Einsteina bozonów NG. Parametr porządku może przyjmować różne wartości w danym zakresie i jest to zależne od temperatury. Powyżej temperatury krytycznej TC zanika i następuje przejście fazowe do stanu symetrycznego z utratą uporządkowanej struktury (przywrócenie symetrii). Zobacz powyżej przykłady wartości TC (w diamentach, magnesach, nadprzewodnikach). QFT pozwala zatem na opis różnych faz, w których można znaleźć system. Te różne fazy charakteryzują się fizycznie różnymi typami zachowań w zależności od różnych wartości parametru porządku i są opisywane przez fizycznie różne przestrzenie stanów układu, tj. jednostkowo nierównoważne reprezentacje kanonicznych relacji komutacyjnych (CCR). W rzeczywistości w QFT istnieje nieskończenie wiele jednostkowo nierównoważnych reprezentacji CCR. Nie istnieją one w QM ze względu na twierdzenie von Neumanna, które stwierdza, że dla układów o skończonej liczbie stopni swobody wszystkie reprezentacje CCR są jednostkowo (a zatem fizycznie) równoważne. Pola z definicji opisują układy o nieskończonej liczbie stopni swobody, zatem twierdzenie von Neumanna nie obowiązuje dla nich. Systemy, które mogą mieć różne fazy fizyczne, należy zatem opisać za pomocą QFT, który może uwzględnić wielość ich faz i przejść między nimi, a nie za pomocą QM.

Model wielu ciał mózgu

Dylemat Lashleya i problemy pojawiające się w badaniu mózgu, mają swoje źródło w ogromnej liczbie składników mózgu na poziomie komórkowym i subkomórkowym oraz w ogromnej złożoności ich organizacji i dynamiki. Z drugiej strony stabilność funkcjonalnej aktywności mózgu jest niezbędna w każdej naszej działalności, a nawet dla naszego przetrwania w świecie. Jak może powstać z niezliczonych składników mózgu? Istnieje około 10¹¹ neuronów z 10¹⁵ synapsami, każdy z nich jest w stanie wyzwolić około 10 impulsów na sekundę, co oznacza około 1016 operacji synaptycznych na sekundę, nie wspominając o komórkach glejowych i fakcie, że wszystko to występuje w kąpieli o 90% większej liczne cząsteczki wody. Każda z tych cząsteczek niesie oscylujący elektryczny pęd dipolowy podlegający nieuniknionym fluktuacjom kwantowym. Jednak całkowita aktywność mózgu wymaga zużycia energii na sekundę rzędu zaledwie 25 W. To śmiesznie mało w porównaniu z mocą niezbędną do symulacji całkiem elementarnych zadań przez jeden z gigantycznych amerykańskich lub europejskich projektów Brain Projects, który jest rzędu 1,5 MW. Jak już wspomniano powyżej, cytując Ricciardiego i Umezawę, należy również wyjaśnić, jak to się dzieje, że na sprawność funkcjonalną mózgu nie ma wpływu nieprawidłowe funkcjonowanie lub nawet utrata pojedynczych neuronów. Aktywność metaboliczna powoduje przemiany chemiczne i wymiany biomolekuł w odstępach czasu rzędu kilku tygodni. Trudno jest wówczas wyjaśnić długi i średni czas życia wspomnień w kategoriach zlokalizowanych układów biomolekuł, ze względu na ich zmiany i odnowę w takim procesie wymiany. Schrödinger zauważył, że „enigmatyczna stabilność biologiczna" żywej materii (ale jak już powiedziano, jego obserwacja może dotyczyć także mózgu) nie da się wytłumaczyć w kategoriach „prawidłowości tylko w średniej" wywodzącego się z „mechanizmów statystycznych". Według niego byłoby to „oczekiwanie fizyka klasycznego", że „nie jest wcale trywialne, ale jest błędne" Wychodząc z tych uwag, w 1967 r. Ricciardi i Umezawa zaobserwował :

[...] wydaje się, że uzyskano bardzo niewiele konkretnych wyników w tym sensie, że wciąż aktualne jest pytanie, w jaki sposób mózg przetwarza informacje otrzymane z zewnątrz i na jakiej logice opierają się operacje wykonywane przez mózg dalekie od zadowalającego rozwiązania. […] W sposób naturalny pojawia się więc jedna możliwość: skoro zwykle ignoruje się mechanizm, według którego mózg dokonuje inteligentnych operacji, [...] można by spróbować podać bardziej ogólny opis dynamiki mózgu; [...] z fenomenologicznego punktu widzenia silnie sugeruje model kwantowy. Innymi słowy, można próbować szukać specyficznych mechanizmów dynamicznych (znanych już w fizyce o wielu stopniach swobody), które mogą zaspokoić istotne wymagania obserwowanego funkcjonowania mózgu.

W wielociałowym (kwantowym) modelu mózgu sformułowanym przez Ricciardiego i Umezawę (RU) zakłada się, że za łamanie symetrii odpowiedzialny jest bodziec zewnętrzny odbierany przez mózg. Zakłada się, że gęstość kwantów korelacji NG generowanych w procesie łamania (patrz rozdział 3) jest indeksem lub kodem dystynktywnym pamięci związanej z bodźcem zewnętrznym, który spowodował złamanie symetrii. To samo dzieje się w rozpraszającym modelu kwantowym, który zostanie omówiony poniżej. W obu modelach symetrię załamaną pod wpływem bodźców zewnętrznych utożsamiono z symetrią rotacyjną (sferyczną) molekularnych dipoli elektrycznych.Należy podkreślić, że w modelu RU (oraz w modelu dyssypatywnym) neurony i inne jednostki komórkowe są układami klasycznymi. Zmienne kwantowe to mody wibracyjne dipoli elektrycznych matrycy wodnej i innych obecnych biomolekuł. Korelacje NG dalekiego zasięgu między nimi sprzyjają i podtrzymują składanie i demontaż oscylujących domen populacji neuronowych. Należy zauważyć, że bodziec nie wpływa na wewnętrzną dynamikę mózgu. Wywołuje jedynie spontaniczne załamanie symetrii rotacyjnej dipoli cząsteczek wody. Wewnętrzna dynamika przebiega wówczas w oparciu o właściwości fizyczne i chemiczne mózgu, niezależnie od bodźca. Aspekt ten może mieć jednoznaczną i bezpośrednią weryfikację w laboratorium,

a jego opis stanowi wyróżniającą zaletę kwantowego modelu mózgu. Tłumaczy to także fakt, że bodziec zewnętrzny, nawet odmienny od tego, który pierwotnie indukował proces zapamiętywania, może pobudzić przywołanie wcześniej zapisanego wspomnienia. Wyjaśnia to w sposób dynamiczny powszechnie doświadczane zjawisko przywoływania wspomnień w warunkach percepcyjnych, które również bardzo różnią się od tych, w których zostało to po raz pierwszy zapamiętane. Mamy tutaj na myśli bodźce normalne lub „słabe", a nie silnie stresujące, takie jak wstrząs (lub także elektrowstrząs), który może zniewolić funkcjonalność mózgu. Chociaż bodziec może być dość słaby, aby wywołać SBS, musi być „w fazie" z dynamiką mózgu. W modelu RU i jego późniejszych opracowaniach [45, 46] zapis pamięci wywołanej bodźcem został anulowany przez zapis innej pamięci wywołanej kolejnym bodźcem. To nadrukowanie pamięci zminimalizowało pojemność pamięci modelu. Dla uproszczenia w modelu nie uwzględniono faktu, że mózg jest systemem pozostającym w ciągłej, nieuniknionej interakcji z otoczeniem, wewnętrznie na nie otwartym. „Zamknięcie" mózgu oznacza uszkodzenie jego funkcjonalności, co można zaobserwować w eksperymentach zmuszających osobę do izolacji. Model mózgu RU uantum został zatem zmodyfikowany tak, aby uwzględnić „otwartość" jego dynamiki. Doprowadziło to do sformułowania rozpraszającego modelu kwantowego mózgu.

Rozpraszający model kwantowy mózgu

Żadna próba opisu mózgu nie może ignorować ciągłej i wzajemnej wymiany energii pomiędzy mózgiem a otoczeniem. Znajomość szczegółów biomolekularnych i komórkowych jest fundamentalna, ale wyraźnie niewystarczająca do opisu aktywności mózgu. Samo to nie może uwzględniać właściwości systemu jako otwartego systemu. Jak już widzieliśmy, szczegółowe badanie elementów elementarnych jest konieczne, ale niewystarczające. Musi być uzupełniona znajomością dynamiki rządzącej zbiorem elementów elementarnych. W przypadku mózgu jest to rozpraszająca forma dynamiki. Prowadzi nas to do sformułowania rozpraszającego modelu kwantowego mózgu. Fizyczna potrzeba uwzględnienia mózgu i jednocześnie jego otoczenia przekłada się na matematyczną potrzebę „podwojenia systemu": mamy system mózgowy i system środowiska. Tę ostatnią, której nie da się wyeliminować, można schematycznie przedstawić jako zbiornik, z którego pochodzi wszystko, co mózg wchłania i do którego wlewa się wszystko, co mózg wypuszcza. Ogólny układ mózg–środowisko jest układem zamkniętym, w którym przepływ energii na granicy mózg–środowisko jest doskonale zrównoważony. Z punktu widzenia równoważenia przepływów środowisko opisuje się zatem w taki sam sposób, w jaki opisuje się mózg, pod warunkiem, że „napływ" zamienia się na „wypływ" i odwrotnie, co uzyskuje się algebraicznie przez zmianę znak zmiennej czasu: środowisko jest zatem „odwróconą w czasie" kopią mózgu. Oczywiście interakcja mózgu z otoczeniem jest bardzo złożona i wymaga wiedzy i detekcji ogromnej liczby parametrów. Jeśli jednak ograniczymy się do rozważenia jedynie bilansu przepływów energii, opis środowiska jako odwróconej w czasie kopii mózgu jest matematycznie poprawny i wystarczający. W tym opisie środowisko jest zatem w rzeczywistości „kopią", sobowtórem mózgu. Wzajemne oddziaływanie liniowości i nieliniowości odgrywa interesującą rolę w modelu rozpraszającym. Przejścia fazowe pomiędzy różnymi reprezentacjami (fazami) zachodzą w nieliniowym reżimie dynamicznym (krytyczność). SBS implikuje dynamiczną nieliniowość, w wyniku której powstaje kondensacja bozonów i stany spójne. Liniowość utrzymuje się w każdej fazie. Taka zależność liniowości i nieliniowości jest zgodna z obserwacjami wskazującymi na współistnienie modów falowych i impulsowych. Aktywność impulsową można zaobserwować w eksperymentach opartych na odpowiedzi liniowej. Z drugiej strony ich zsynchronizowane wzorce AM wykazują log-log gęstość mocy w funkcji rozkładów częstotliwości, tj. dynamikę bezskalową (samopodobny fraktal) wymagającą spójności wynikającą z dynamiki nieliniowej. Samopodobne właściwości fraktalne są rzeczywiście izomorficzne ze stanami spójnymi, co jest zgodne z leżącą u podstaw spójną dynamiką wielu ciał modelu rozpraszającego. W dynamicznej ewolucji mózgu („funkcjonowaniu") występują zmiany w przepływach wymienianych z otoczeniem, dlatego też stan mózgu musi być stale aktualizowany, ale

także opis stanu środowiska musi być aktualizowany. zrównoważyć przepływy energii. W stanach pamięci, które są stanami spójnymi dwumodowymi, tryby mózgu są trwale splątane z modami podwójnymi (trybami środowiska). Istnieje zatem ciągła „wzajemna aktualizacja", proces wzajemnej reakcji zwrotnej, „dialog" między mózgiem a jego sobowtórem, nigdy nie jest to monolog, nigdy nierozwiązywalny. Czasami w konflikcie pomiędzy jaźnią a Sobowtórem rozwija się dynamika poznania, zrozumienia, odczuwania i życia. Wzajemne oddziaływanie na siebie wymaga ciągłej aktualizacji ich relacji. Każdy z nich jest wystawiony na spojrzenie drugiego. Należy zauważyć, że relacja splątania implikowana jest przez korelację fazową pomiędzy modami, która nie wymaga wymiany komunikatora ani informacji, a zatem może być ustalona natychmiastowo, bez naruszania postulatów teorii względności. Korelacja nie jest zatem interakcją, która wymagałaby wymiany posłańca, którego prędkość nie mogłaby być wówczas większa od prędkości światła. Wracając do dialogu pomiędzy jaźnią a Sobowtórem, to właśnie w tym „entre-deux" akt świadomości ma swój początek [8, 9]; podsumowuje w sobie doświadczenie zgromadzone w przeszłości, ale składa się tylko z teraźniejszości. W tej perspektywie mózg wydaje się być „rozszerzony" w swojej funkcjonalności, poza granice swojej konfiguracji anatomicznej. Świadomość rozszerza się na środowisko, w którym zanurzony jest mózg. Należy podkreślić, że relacja z Sobowtórem jest relacją dynamiczną, a nie narcystycznym odbiciem. W modelu rozpraszającym nie ma takiego odbicia lustrzanego. Jak zauważa Desideri, nawiązując do niektórych dyskusji na temat neuronów lustrzanych, odbicie lustrzane jest statyczne i nie stanowi okazji do uczenia się, ponieważ obserwowane działanie i wykonywane działanie są strukturalnie równoważne. To, co obserwuje się w laboratorium i należy do naszego powszechnego doświadczenia, to właściwość mózgu do gromadzenia doświadczeń i budowania wiedzy, czyli uczenia się, jak „maksymalnie chwycić" świat. W tym celu nie wystarczy kopia, proste odbicie lustrzane; potrzebny jest zabieg twórczy, mimesis w sensie Poetyki Arystotelesa, która – jak podkreśla Desideri – dotyczy tego, co możliwe, a nie tego, co po prostu się wydarza. Potrzebujemy takiej ilości wyobrażeniowej nieokreśloności, która umożliwia uczenie się, a także zmianę obserwowanego modelu działania [60]. Godne uwagi jest to, że model dyssypatywny pozwala na taki stopień swobody i że proces uczenia się wynikający z dialogu z Sobowtórem jest formalnie powiązany z minimalizacją darmowej energii systemu. Zauważam, że zrównoważenie przychodzących i wychodzących przepływów energii jest równoznaczne z ustawieniem ich różnicy na zero. Charakteryzuje to stan równowagi całego układu mózg-środowisko. Jednakże ustawienie różnicy między dwiema wielkościami na zero pozostawia je całkowicie nieokreślone. Operacja równoważenia pozwala zatem na nieskończoną serię par stanów odpowiednio mózgu i środowiska, dla których taka różnica wynosi zero. Każdemu z tych stanów mózgu (oraz odpowiadającym im stanom środowiska w każdej z par) odpowiada inna wartość gęstości skondensowanych kwantów. Każdą z tych gęstości można uznać za indeks lub kod pamięci. Można wykazać, że stany o różnych gęstościach są względem siebie ortogonalne (jednostkowo nierównoważne, patrz rozdział 3), a zatem nie zakłócają się wzajemnie. Wspomnienia są w ten sposób chronione przed wzajemnym „zamieszaniem". Widzimy, że jednostkowa nierównoważność reprezentacji QFT odgrywa zatem kluczową rolę w procesie zapisywania w pamięci. Co więcej, ich nieskończona liczba gwarantuje dużą pojemność pamięci. W rezultacie, dzięki rozproszeniu, możemy mieć nieskończenie wiele niezakłócających stanów pamięci. Rozpraszanie rozwiązuje problem pojemności pamięci. Ogromna pojemność pamięci jest konsekwencją faktu, że mózg jest systemem rozpraszającym, otwartym.

Chaotyczne trajektorie w krajobrazie atraktorów. Błędy i przemyślenia

Z tego, co powiedziano powyżej, widzimy, że nabycie „nowej" pamięci odpowiada wykorzystaniu jednego z nieskończenie wielu jednostkowo nierównoważnych stanów podstawowych (próżni), do których dostęp ma system mózg-środowisko. Możemy zatem opisać zbiór stanów pamięci (lub „przestrzeń pamięci") jako zbiór takich spójnych stanów podstawowych, z których każdy jest

oznaczony kodem konkretnej pamięci. Są to stany o minimalnej energii, ponieważ w nich różnica między przychodzącymi i wychodzącymi przepływami energii wynosi zero, jak wspomniano powyżej. Co więcej, są to także stany minimalnej energii swobodnej. Są to więc stany, do których system „zmierza" w swojej ewolucji, jak do „atraktorów". Zbiór stanów pamięci przedstawia zatem „krajobraz atraktorów". Godne uwagi jest to, że ścisła matematyczna jednostkowa nierównoważność reprezentacji zostaje wygładzona w realistycznych sytuacjach z powodu defektów, zanieczyszczeń i efektów powierzchniowych. Taka „niedoskonałość" jest jak najbardziej pożądana, ponieważ pozwala na ewolucję wspomnień w czasie, a tym samym na możliwość „zapominania" i wspomnień o różnym czasie życia, a więc wspomnień krótkotrwałych i długoterminowych. Co więcej, umożliwia także przejścia, czyli ścieżki, trajektorie poprzez wspomnienia (poprzez stany pamięci), tak że między atraktorami mogą powstawać korelacje, gdy mózg przechodzi od atraktora do atraktora; rzeczywiście wzdłuż trajektorii w krajobrazie atraktorów, zamieszkując mniej więcej przez długi czas w każdym z nich, nigdy w normalnych (zdrowotnych) warunkach nie będąc tam uwięzionym. Wzdłuż każdej trajektorii energia swobodna układu jest minimalna. W modelu dyssypatywnym kluczową rolę odgrywa energia swobodna i jej minimalizacja. To faktycznie kontroluje gęstość kondensatu oznaczającego każdy stan pamięci. Należy podkreślić, że nabycie nowej pamięci wiąże się nie tylko z dodaniem nowego atraktora do krajobrazu atraktorów, ale z reorganizacją całego krajobrazu, a co za tym idzie, jego całkowitą aktualizacją w świetle nowego nabytku. Na tym procesie polega kontekstualizacja nowego nabytku i wyłonienie się jego znaczenia, które nigdy nie należy do bodźca percepcyjnego (do wejścia). Wpisuje się w kontekst przeprojektowanego krajobrazu atraktorów, zawsze w całości nowego. Treść znaczeniowa korelacji w przestrzeni atraktorów nie jest zatem nigdy ostateczna. Znaczenia zawsze można zaktualizować, lepiej zrozumieć lub całkowicie zmienić. Zawsze są poddawani próbom. Pojawia się zatem wymiar nowości, zaskoczenia, a nawet zdumienia, związany z nagłym zobaczeniem czegoś nieoczekiwanego [59, 64]. W tym odmiennym ujęciu [65] należy szukać genezy wyobraźni i jej roli w wyznaczaniu różnych trajektorii w przestrzeni atraktorów. Daleko nam do prostego odbicia lustrzanego. Relacja mózgu ze światem jest całkowicie dynamiczna. Proces kontekstualizacji, w którym mózg kwestionuje całą swoją historię doświadczeń, stanowi jedną z najbardziej charakterystycznych cech rozpraszającego kwantowo modelu mózgu. Wiernie opisuje laboratoryjne obserwacje, podczas których badany podmiot, zwierzę czy człowiek, reaguje na sytuacje, w których znajduje się poddawany procesowi abstrakcji (lub egzemplifikacji niezbędnej do zbudowania nowego atraktora, czyli równoważenia przepływów) i generalizacji (czyli tworzenia kategorii w ustalaniu korelacji w krajobrazie atraktorów). W ten sposób przepływ informacji wymienianych w relacji ze światem staje się wiedzą [31, 59, 66], a pamięć staje się pamięcią znaczeń, a nie pamięcią informacji. Każdy akt rozpoznania krajobrazu atraktorów reprezentuje akt wiedzy intuicyjnej, rozpoznanie trybu zbiorowego, niepodzielnego na racjonalne kroki, możliwego do pomyślenia, ale „nieobliczeniowego" i nieprzetłumaczalnego na logiczne ramy języka . Ta cecha aktywności mózgu jest być może zgodna ze stwierdzeniem von Neumanna :

Potrzebujemy wyjątkowej precyzji numerycznej w wielu logicznych krokach, aby osiągnąć to, czego mózg dokonuje w zaledwie kilku krótkich krokach.

Można wykazać, że trajektorie w krajobrazie atraktorów, od pamięci do pamięci, są trajektoriami klasycznymi, chociaż „łączą" stany kwantowe. Co więcej, są to trajektorie chaotyczne, to znaczy nie są okresowe (trajektoria nigdy się nie przecina) i trajektorie o różnych warunkach początkowych nigdy się nie przecinają; raczej są (wykładniczo) rozbieżne. Formalnie można wykazać, że chaotyczność trajektorii ma swoje źródło w kwantowej naturze stanów pamięci . Rolę chaosu opisaną w modelu dyssypatywnym potwierdzają obserwacje laboratoryjne. Freeman podkreślił, że :

Chaos jest widoczny w tendencji ogromnych zbiorów neuronów do nagłego i jednoczesnego przechodzenia od jednego złożonego wzorca aktywności do drugiego w odpowiedzi na najmniejszy sygnał wejściowy […] Ta zmienność jest główną cechą wielu chaotycznych systemów […] W rzeczywistości proponujemy, że jest to właśnie ta właściwość, która umożliwia percepcję. Spekulujemy również, że chaos leży u podstaw zdolności mózgu do elastycznego reagowania na świat zewnętrzny i generowania nowych wzorców aktywności, w tym tych, które są odbierane jako świeże pomysły.

Warto zauważyć, że chaotyczna charakterystyka trajektorii w krajobrazie atraktorów sprzyja wysokiej rozdzielczości percepcyjnej. W rzeczywistości minimalne różnice w percepcji (takie, jakie mogą wystąpić w rozpoznawaniu obrazów, zapachów, smaków itp.) są rozpoznawane w krótkim czasie z powodu rozbieżności (chaotycznych) trajektorii. W rzeczywistości rozbieżne trajektorie można łatwo rozpoznać jako różne. Niewielkie różnice w warunkach początkowych przy braku chaotyczności generowałyby niezróżnicowane trajektorie, a rozpoznanie takich różnic byłoby znacznie trudniejsze. W swojej czasowej ewolucji mózg jawi się zatem jako system, który „żyje" na wielu mikroskopijnych konfiguracjach opisanych przez minimalne stany energetyczne odpowiadające różnym wspomnieniom, przechodząc od konfiguracji do konfiguracji (od pamięci do pamięci) swoimi ścieżkami w krajobrazie atraktory (krytyka dynamiki przejść fazowych). Nawet słabe zaburzenie zewnętrzne (słaby bodziec) może wywołać przejścia przez te stany o najmniejszej energii. W ten sposób sporadyczne (losowe) zaburzenia odgrywają ważną rolę w złożonej aktywności mózgu. Z drugiej strony wykazano związek pomiędzy wspomnianym podwojeniem stopni swobody a szumem kwantowym stanów podstawowych. Niezerowe mody podwójne mogą rzeczywiście umożliwiać efekty kwantowe wynikające z urojonej części działania, które nie pojawiałyby się na poziomie klasycznym. Jak właśnie zaobserwowano, przewidywana przez model dyssypatywny rola chaosu i szumu znajduje potwierdzenie w obserwacjach laboratoryjnych, ze szczególnym uwzględnieniem stanu spoczynkowego mózgu, którego dynamika wykazuje samopodobieństwo fraktalne . W warunkach niskiego stopnia otwarcia mózgu na otoczenie, np. podczas snu, pod wpływem substancji psychoaktywnych, podczas medytacji lub w innych stanach obniżonej percepcji zmysłowej, krytyczność dynamiki wzrasta . Chaotyczne trajektorie w przestrzeni pamięci przedstawiają następnie wizualne doświadczenia mózgowe zachodzące w takich warunkach. Rzeczywiście, doświadczenia te często charakteryzują się filmowymi sekwencjami obrazów, z nagłymi przejściami od jednego wzorca obrazu do drugiego. Prawdziwość i realizm odczuwany w tych wizualnych doświadczeniach można omówić w kategoriach algebry podwojenia stopni swobody. W stanach niskiej otwartości mózgu „ja" prawie nie dostrzega sobowtóru jako odrębnego, a ich niemal całkowite dopasowanie wprowadza swego rodzaju „funkcję oceny prawdy", dzięki której prawdziwość i realizm doświadczenia wizualnego jest potwierdzana przez natychmiastowe i jednoznaczną informację zwrotną . Silny wpływ na trajektorie ze względu na minimalne zmiany ich warunków początkowych skłania do rozważenia roli „wątpliwości" w modelu dyssypatywnym. Dialog z Sobowtórem opiera się na ciągłej restrukturyzacji krajobrazu atraktorów, a to z kolei może wywołać, w procesie kwestionowania samego siebie i słuchania [73], słabe zakłócenia w warunkach początkowych trajektorii, co w konsekwencji objawia się ich rozbieżność. W tym procesie można skłonić aktywność mózgu do opuszczenia innych ścieżek lub ucieczki z pułapki („fiksacji") w tym lub innym atraktorze. Wątpliwości można dobrze rozumieć jako błąkanie się po krajobrazie atraktorów spowodowane niepewnością związaną z jego ciągłym przerysowywaniem, wywołane uwodzeniem nowych perspektyw otwieranych przez rozwój nowej trajektorii, kwestionujących pewność nabytą w poprzednich doświadczeniach percepcyjnych. Jest to więc właśnie to wędrowanie (errare) w poszukiwaniu możliwego, niezadowolenie z tego, co się po prostu zdarza [60], charakterystyczna cecha aktywności mózgu, myślenia (pensare). Dlatego mózg nie jest głupią gwiazdą. Zachowuje się raczej jak „urządzenie błędne", „maszyna popełniająca błędy" .

Sztuczna inteligencja i maszyny popełniające błędy. Spartakus

Maszyna to z definicji i konstrukcji urządzenie, które wykonuje szereg czasowo powiązanych operacji w sposób ściśle przewidywalny, niczym w łańcuchu logicznych kroków. Maszyna, która nie przejdzie tak określonego łańcucha etapów, jest zatem maszyną, która nie działa prawidłowo i wymaga naprawy lub wymiany. Wręcz przeciwnie, jak widzieliśmy, aktywność mózgu przebiega etapami które niekoniecznie należą do jednoznacznie określonego łańcucha etapów. Być może moglibyśmy wtedy pomyśleć o mózgu jako o „maszynie popełniającej błędy". Nasz wielki przywilej możliwości popełniania błędów ma swoje korzenie w tym, że błądzimy chaotycznymi trajektoriami w krajobrazie atraktorów, z których wyłania się nieprzewidywalność ruchów świadomości, ich niewierność z góry ustalonego schematu, ich niezbywalnej podmiotowości i całkowitej autonomii. Jak wspomniano powyżej, w modelu dyssypatywnym przyczyną chaotyczności trajektorii jest kwantowa natura dynamiki. W badaniach nad sztuczną inteligencją problemem zbudowania „inteligentnej" maszyny może więc być po prostu problem zbudowania urządzenia zdolnego do popełniania błędów, „urządzenia niekonsekwentnego". Nie jest to maszyna „nieprawidłowo działająca" lub „niesprawna". Jego główna użyteczność nie polega na przewidywalnym zachowaniu, ale raczej na „nowościach" pojawiających się w jego zachowaniu. Błąd lub pomyłka musi jednak mieć charakter „wyjątkowy" w odniesieniu do normalnego lub „poprawnego" zachowania się urządzenia (z definicji jego głupiego lub nudnego zachowania, głupia gwiazda Arystotelesa). „Nowe" lub „inteligentne" rozwiązanie danego problemu proponowane przez wadliwe urządzenie nie należy do listy znanych możliwych rozwiązań ewentualnych problemów zawartych w podstawowej instrukcji urządzenia. To przewidywalne rozwiązania np. automatycznego pilota samolotu AI. Taki automatyczny pilot rzeczywiście musi zostać zastąpiony przez człowieka w przypadku nieprzewidzianej sytuacji awaryjnej, wymagającej rozwiązania, którego „nie ma na liście" automatycznego pilota. W powyższych uwagach odniesienie do błędów nie jest rozumiane w sensie błędów związanych z obserwatorem, ale do błędów wynikających z zachowania błędnych urządzeń, a nie w odniesieniu do oczekiwań obserwatora. W przypadku błędów związanych z obserwatorem wiadomo, że problem „dobra i zła", „prawdy i fałszu" polega w dużej mierze na wyborze modelu przyjętego przez obserwatora. W ramach przyjętego modelu teoria błędów pomaga w ocenie błędów, uwzględniając także możliwość ingerencji obserwatora w pomiar i badane zjawisko. Błędy obserwatora w swym odbieganiu od oczekiwań obserwatora mogą mieć charakter „odstępstwa" od prawidłowego zachowania. Można nawet zdefiniować trywialną maszynę popełniającą błędy, czyli maszynę robiącą coś innego, niż oczekuje pojedynczy obserwator, oraz nietrywialną maszynę popełniającą błędy, robiącą coś innego niż tego oczekuje jakikolwiek obserwator. W przypadku błędów niezwiązanych z obserwatorem nieprzewidywalność błędu oznacza, że „nie można go oczekiwać ani nieoczekiwać w żadnym danym kontekście". Nie jest to zatem ani zaprzeczenie, ani odchylenie. Jest „bezpłatny", a nie „wyprowadzalny", co wskazuje na „nieobliczeniową" aktywność urządzenia. Interesujące jest to, jak pewna postawa konserwatywna, myląca nowość z odchyleniem, odbiera nowość jako atak na własny model (status quo), a nie jako dodatek do niego, co może skutkować wzrostem i wzmocnieniem samego modelu. Nie wyklucza się jednak drugiej alternatywy, a mianowicie tego, że czysty konserwatysta, nawet jeśli nie myli nowości z odchyleniem od normy, jest przeciwny jakiejkolwiek możliwej nowości, aby być po bezpiecznej stronie i nie ryzykować, że nowość unieważni wcześniej istniejącą model w całości lub w części. Podsumowując, widzimy, że samoistnie błędne urządzenie, wytwarzające błędy niezwiązane z obserwatorem, jak opisano, „nie może być maszyną Turinga, czyli algorytmem generującym błędy. Rzeczywiście nie jest możliwe zaprojektowanie algorytmu, który robi tylko oczekiwany wynik, nawet jeśli taki wynik zostanie zdefiniowany jako błędny w odniesieniu do pewnych kryteriów" . Powstaje zatem pytanie, czy możliwe jest zaprojektowanie urządzeń błędnych niezwiązanych z obserwatorem. Co ciekawe, skoro błąd niezwiązany z obserwatorem jest nowością, „pojawieniem się" w zachowaniu systemu, to pytanie takie można powiązać także z problemem projektowania „pojawienia się", rozpatrywanego jako możliwy błąd pojawiający się w procesach popełniania błędów . Alternatywą dla

wewnętrznie błędnego urządzenia byłaby po prostu proteza, która pomogłaby nam w przypadku niektórych naszych braków fizycznych lub behawioralnych albo poprawiła lub wzmocniła nasze ograniczone zdolności (sztuczne ramię, prawdopodobnie kontrolowane poprzez połączenia z naszymi nerwami, duża pojemność pamięci z bardzo krótkim czasem potrzebnym na uporządkowanie jednego z nich, komputer, automatyczny pilot, telefon komórkowy, określany jako inteligentny z subtelnym poczuciem humoru, robot itp.), czyli mniej więcej to, co AI zapewnia obecnie. Być może program budowy świadomego sztucznego urządzenia przechodzi przez konstrukcję wewnętrznego urządzenia błędnego. Jeśli kiedykolwiek będzie możliwe zbudowanie urządzenia wyposażonego w świadomość, musi ono posiadać wszystkie najlepsze właściwości charakteryzujące człowieka: nieprzewidywalność jego zachowań, zdolność uczenia się, ale także niewierność, nieuniknione zaangażowanie w świat. świata i jego niezbywalną wolność. I on/ona musi nazywać się Spartakus

Uwagi końcowe

Podsumowując, widzieliśmy, że SBS implikuje dynamiczne generowanie fal korelacyjnych dalekiego zasięgu pomiędzy elementarnymi składnikami układu oraz że powiązane kwanty, czyli mody bozonów NG, są bezmasowe. Ich spójna kondensacja w stanie podstawowym układu odpowiada za obserwowalny tam porządek. Symetria odpowiada niezmienności obserwowalnych właściwości systemu; stany układu przed i po transformacji symetrii są „nierozróżnialne". Porządek wynikający z załamania symetrii odpowiada natomiast możliwości rozróżnienia stanu układu przed i po przekształceniu symetrii; W ten sposób powstaje „różnorodność". Scenariusz powstały poprzez SBS w QFT charakteryzuje się ogromnym bogactwem form. Prowadzące do nich procesy dynamiczne, które możemy nazwać procesami morfogenetycznymi, w rzeczywistości opisują dynamiczne generowanie wielu różnych obserwowalnych przejawów tych samych podstawowych równań dynamicznych, „narastanie różnic" w otaczającym nas świecie. To bogactwo różnorodności, chwała Babel. Systemy rozpraszające nie są zamknięte w sobie. Wymieniają energię, pęd, masę itp. Z otoczeniem. Następnie w każdym z nich zostaje złamana niezmienność translacyjna w czasie, co oznacza, że początek osi czasu nie może być swobodnie przełożony, przez co czas nie jest już zmienną fikcyjną. Systemy rozpraszające są systemami starzejącymi się. Historia ma swoje korzenie. Model mózgu wielu ciał i rozpraszający model kwantowy mózgu powstały w wyniku zastosowania formalizmu QFT SBS do odpowiedzi na niektóre otwarte pytania w neurologii, jak opisano w powyższych sekcjach. Podwojenie stopni swobody wymaganych przez formalizm matematyczny dla systemów rozpraszających doprowadziło do wprowadzenia „lustrzanego w czasie" obrazu siebie, czyli Double. W dialogu z Sobowtórem wiedza budowana jest na bazie doświadczeń zgromadzonych w przeszłych percepcjach. Z tego wyłania się perspektywa, czyli wizja świata. Intencjonalność, która determinuje nasze działanie, ma swoje korzenie w aktualizowaniu nigdy ostatecznej równowagi z otaczającym nas światem, generując znaczenia i znaczące relacje z nim. Może się zdarzyć, że urzeczywistni się „idealna wymiana pomiędzy wnętrzem i zewnętrzem", „korzystne połączenie" pomiędzy „ja" a obiektem. Według Desideriego jest to przeżycie estetyczne. Estetyczne nie jest zatem doświadczeniem konkretnym, ani byle jakim, ale „jako wymiar przenikający całe pole naszego doświadczenia (i fakturę percepcyjną, która konfiguruje «krajobraz»)". Rozpoznanie takiego doświadczenia determinuje sąd estetyczny, który „zawsze dotyczy pierwszej osoby". Efektem jest „nowe spojrzenie na świat", który nie jest obcy, ale konkurent w wymiarze poznawczym. Rozbieżność trajektorii w krajobrazie atraktorów gwarantuje bowiem, że przeżycie estetyczne jest zawsze nowe i wywrotowe w porównaniu z konsolidacją już zbadanych krajobrazów. Orientacja, którą wyraża, „zawsze oczekuje odnowy", ponieważ równoważenie przepływów, którego jest wyrazem, ma charakter dynamiczny, nigdy ostateczny. Emocjonalna reakcja na doświadczenie estetyczne ma zatem wartość performatywną w łuku intencjonalnym. Doznanie estetyczne jest więc cechą charakterystyczną dynamiki mózgu. Chaotyczne klasyczne trajektorie przechodzące przez stany pamięci charakteryzują dynamikę mózgu, oferując mózgowi zdolność do zapewniania

nieprzewidywalnych zachowań i odpowiedzi na doświadczenia percepcyjne. To prowadzi nas do przedstawienia mózgu jako urządzenia z natury błędnego, zdolnego do wykonywania kroków, które nie są powiązane ścisłym, jednoznacznie określonym następstwem. Jak wspomniał i zauważył Freeman, chaos rzeczywiście odgrywa istotną rolę w aktywności mózgu. Te właściwości aktywności funkcjonalnej mózgu wydają się trudne do modelowania w ramach programu badawczego AI. Trudno uzasadnić tę samą motywację dotyczącą projektu wewnętrznie błędnego urządzenia AI, ponieważ takie urządzenie jest w pewnym sensie całkowitym przeciwieństwem posłusznej, lojalnej maszyny wykorzystywanej w rzeczywistych projektach AI. Od robota doskonałego wymaga się wierności i niezawodności w realizowaniu zadań społecznych, przemysłowych czy wojskowych uzasadniających nakłady finansowe włożone w jego budowę. Co więcej, musi mieć stosunkowo krótki czas życia, aby nie nasycić rynku (kluczowy wymóg stosowany już w smartfonach, samochodach, telewizorach, zmywarkach itp.). Po zbudowaniu (i sprzedaniu) jego wartość komercyjna musi szybko dążyć do zera (i rzeczywiście tak jest). Te realistyczne elementy „kruchości" nieodłącznie związane z urządzeniami AI utrudniają im bycie czymś więcej niż protetyką w przypadku niektórych naszych własnych braków fizycznych lub niepełnosprawności. Niestety dzisiejsze projekty AI nadal ograniczają się do projektowania „głupich gwiazd". Chciałbym zakończyć obserwacją na granicy mózgu i środowiska. Kiedy środowisko składa się z innych, pytanie brzmi: gdzie kończy się „ja", a gdzie zaczynają „inni"? Pytanie staje się jeszcze bardziej radykalne, gdy zadawane jest przez grupy ludzi, które czują, że mają „silną tożsamość": gdzie kończy się „my", a zaczyna „oni"? Czy potrafimy wyobrazić sobie świat bez innych? Jedna z możliwych odpowiedzi brzmi: „Nie. Jesteśmy wszyscy razem, my sami jesteśmy innymi". Albo odpowiedź odwrotna: „Nie oni, tylko my. Jesteśmy pierwsi, oni są inni od nas". Ale to drugie nie jest zgodne z otwartością dynamiki mózgu. Zamknięcie, które proponuje, jest równoznaczne z samobójstwem. Inni także są częścią naszego sobowtóra, a mianowicie nas samych. „Ich eliminacja" byłaby samoeliminacją. Należymy do siebie; całe bogactwo wyobraźni i kreatywności dialogu z Sobowtórem wkracza w naszą wzajemną relację. Mózg ma nieodłączny wymiar społeczny. Na koniec być może interesujące będzie zacytowanie fragmentu z książki „Borges i ja", świadczącego o szerokim imaginacyjnym horyzoncie dialogu pomiędzy jaźnią a jej sobowtórem: Ten drugi, ten zwany Borgesem, jest tym, co przydarza się […] Byłoby przesadą stwierdzenie, że nasze stosunki są wrogie; Żyję, pozwalam sobie żyć dalej, aby Borges mógł wymyślać swoją literaturę i ta literatura mnie usprawiedliwia […] Poza tym jestem skazany na śmierć definitywną i tylko jakaś chwila mnie może go przeżyć […] Spinoza wiedział o tym wszystkie rzeczy pragną trwać w swoim bycie; kamień wiecznie chce być kamieniem, a tygrys tygrysem. Pozostanę w Borges, a nie w sobie (jeśli prawdą jest, że jestem kimś) […] Przed laty próbowałam się od niego uwolnić i przeszłam od mitologii przedmieść do gier z czasem i nieskończonością, ale te gry należą teraz do Borgesa i będę musiał wyobrazić sobie inne rzeczy. W ten sposób moje życie jest ucieczką i tracę wszystko i wszystko należy do zapomnienia lub do niego. Nie wiem, który z nas napisał tę stronę.

Trudny problem i wolna wola: podejście informacyjno-teoretyczne

Jesteśmy czymś, z czego powstają sny, a nasze małe życie dopełnia sen. William Szekspir

Streszczenie Badamy określone twierdzenia teoretyczne na temat świadomości, zaczynając od nieredukcyjnego psychoinformacyjnego rozwiązania trudnego problemu Davida Chalmersa, opartego na hipotezie, że podstawową właściwością „informacji" jest jej doświadczanie przez wspierający „system". Rodzaj informacji zaangażowanych w świadomość musi mieć charakter kwantowy z wielu powodów, w tym z powodu jej wewnętrznej prywatności i mocy budowania myśli poprzez splątanie stanów qualiów. W rezultacie dochodzimy do panpsychizmu opartego na informacji kwantowej, w którym fizyka klasyczna superweniuje na fizyce kwantowej, fizyka kwantowa superweniuje na informacji kwantowej, a informacja kwantowa superweniuje na świadomości. Następnie

argumentujemy, że wewnętrznie doświadczany stan kwantowy, ponieważ odpowiada określonemu doświadczeniu – a nie przypadkowemu wyborowi – musi być czysty i nazywamy go ontycznym. Należy to odróżnić od stanu przewidywalnego z zewnątrz (tj. stan opisujący poznanie doświadczenia z punktu widzenia obserwatora zewnętrznego), który nazywamy epistemicznym i na ogół ma charakter mieszany. Czystość stanu bytowego wymaga ewolucji zachowującej czystość, czyli tzw. atomowej operacji kwantowej. To drugie rozwiązanie ma charakter ogólnie probabilistyczny, a jego konkretny wynik interpretuje się jako wolną wolę, która jest nieprzewidywalna nawet w zasadzie, ponieważ prawdopodobieństwa kwantowego nie można interpretować jako braku wiedzy. Widzimy także, jak ta sama czystość stanu i ewolucji pozwala rozwiązać dobrze znany złożony problem panpsychizmu. Ewolucja stanu kwantowego stanowi krótkotrwały bufor doświadczenia i sama zawiera transfery informacji z kwantu do klasycznego i z klasycznego do kwantowego. Z drugiej strony pamięć długoterminowa jest klasyczna i wymaga zapamiętywania i przypominania procesów, które są odpowiednio kwantowe i klasyczne. W procesach takich można wykorzystać wielokrotne kopie stanu doświadczanego, przygotowane na nowo z „uwagą", co pozwala na lepszą jakość klasycznego przechowywania. Wreszcie badamy możliwość eksperymentalnych testów naszej teorii w naukach kognitywistycznych, w tym ocena liczby zaangażowanych kubitów, istnienie komplementarnych obserwabli i naruszenia granic lokalnego realizmu. W dodatkach zwięźle ilustrujemy ramy operacyjnej teorii probabilistycznej (OPT) dla możliwych postkwantowych teorii świadomości, oceniając dogodne czarno- podejście pudełkowe OPT, wraz z jego solidnością metodologiczną w oddzielaniu elementów obiektywnych od teoretycznych, gwarantującą kontrolę eksperymentalną i falsyfikowalność.Na koniec syntetycznie porównujemy matematyczne postulaty i twierdzenia najważniejszych przypadków OPT – tj. teorii klasycznych i kwantowych – aby pomóc czytelnik może lepiej zrozumieć naszą teorię świadomości. Zapis matematyczny znajduje się w przydatnej tabeli w załącznikach.

Panpsychizm kwantowo-informacyjny

W swojej książce The Character of Consciousness [1] David Chalmers stwierdza to, co nazywa trudnym problemem świadomości, a mianowicie kwestią wyjaśniania naszych doświadczeń – zmysłowych, cielesnych, mentalnych i emocjonalnych, w tym dowolnego strumienia myśli. Chalmers zestawia problem trudny z problemami łatwymi, które, jak we wszystkich naukach, można rozwiązać w kategoriach podejścia mechanistycznego, które jest bezużyteczne w przypadku problemu doświadczenia. Rzeczywiście, we wszystkich naukach zawsze szukamy wyjaśnień w kategoriach funkcjonowania, koncepcji całkowicie niezależnej od pojęcia doświadczenia. Chalmers pisze:

Dlaczego pełnieniu tych funkcji towarzyszy doświadczenie? Dlaczego całe to przetwarzanie informacji nie odbywa się „w ciemności" i pozbawione jakichkolwiek wewnętrznych odczuć? Istnieje luka wyjaśniająca pomiędzy funkcją a doświadczeniem.

Skutecznym paradygmatem pozwalającym uchwycić pojęciową lukę między „doświadczeniem" a „funkcjonowaniem" jest paradygmat zombie, którego behawioralnie nie można odróżnić od świadomej istoty, niemniej jednak nie ma on wewnętrznego doświadczenia. Obecnie istnieją dwie główne linie odpowiedzi na ten trudny problem: (1) pogląd fizykalistyczny – ze świadomością „wyłaniającą się z funkcjonowania", taką jak jakaś biologiczna właściwość życia; (2) pogląd panpsychistyczny – ze świadomością jako podstawową cechą świata, którą posiadają wszystkie byty. Proponuje się tutaj: panpsychizm ze świadomością jako podstawową cechą „informacji" i fizyką superweniującą na informacji. Pomysł, że fizyka jest przejawem czystego przetwarzania informacji, był zdecydowanie propagowany przez Johna Wheelera [3] i Richarda Feynmana, a także kilku innych autorów, w tym Davida Finkelsteina [6], któremu szczególnie podobał się ten pomysł. Jednak dopiero całkiem niedawno został konkretnie ustalony nowy paradygmat informacyjny w fizyce. W ramach tego programu osiągnięto: (1) wyprowadzenie teorii kwantowej jako teorii informacji oraz (2)

wyprowadzenie swobodnej kwantowej teorii pola jako wyłaniającej się z nietrywialnego algorytmu kwantowego z układami przeliczalnymi o minimalnej złożoności algorytmicznej.1 Oprócz takiej wartości metodologicznej, nowe informacyjno-teoretyczne wyprowadzenie kwantowej teorii pola jest szczególnie obiecujące dla ustalenia teoretycznych ram dla kwantowej grawitacji wyłaniającej się z kwantowego przetwarzania informacji, co sugeruje również rola, jaką informacja odgrywa w zasada holograficzna. Podsumowując, świat fizyczny wyłania się z leżącego u jego podstaw algorytmu, a rodzaj informacji przetwarzanej w jego ramach ma charakter kwantowy. Pomysł, że teorię kwantową (QT) można uznać za teorię informacji, jest stosunkowo nowy i wywodzi się z dziedziny informacji kwantowej. Tymczasem to, co nazywamy „teorią informacji", w dużej mierze ewoluowało, od jej początków jako teorii komunikacji [18], w kierunku ogólnej teorii „przetwarzania" informacji, która wcześniej była jedyną domeną informatyki. Co rozumiemy przez „teorię informacji"? Ostatnio, zarówno w fizyce, jak i informatyce (która w międzyczasie była związana z informacją kwantową), w literaturze fizyki wyłoniły się ramy teoretyczne dla wszystkich teorii informacji w ujęciu pojęcia operacyjnej teorii probabilistycznej (OPT), izomorficzna struktura, która pojawiła się w informatyce w kategoriach teorii kategorii. Rzeczywiście, ramy matematyczne teorii informacji są dokładnie takie same jak w przypadku OPT, niezależnie od tego, którą teorię informacji rozważamy – czy to klasyczną, kwantową, czy „postkwantową". Przegląd głównej struktury OPT znajduje się w Dodatku. Wśród teorii informacji znajdują się: teoria klasyczna (CT) odgrywa szczególną rolę.W rzeczywistości, poza tym, że sama jest OPT, CT wchodzi w ramy operacyjne w zakresie obiektywnych wyników teorii, które w przypadku przyczynowych OPT (takich jak QT i CT) mogą zostać wykorzystane do warunkowania wyboru f kolejna transformacja w obrębie zbioru. Oczywiście dzieje się to również w szczególnym przypadku QT. Zatem wystąpienie danego wyniku można uznać za wymianę informacji z kwantowej na klasyczną, natomiast warunkowanie stanowi wymianę informacji z klasycznej na kwantową Dochodzimy do wniosku, że powinniśmy uważać świat fizyczny za wiernie rządzony zarówno przez teorię kwantową, jak i klasyczną, przy czym informacje ulegają transformacji między obydwoma typami. Ten teoretyczny opis rzeczywistości należy skontrastować ze zwykłym poglądem na rzeczywistość jako kwantową, tworząc błędne przekonanie o niewłaściwie umiejscowionej konkretności. Najbardziej pragmatycznym punktem widzenia jest traktowanie QT i CT łącznie jako prawidłowej teorii informacji do opisu rzeczywistości, bez ponoszenia żadnego logicznego paradoksu. Wykorzystamy ten pomysł w dalszej części rozdziału. Ze względu na ukrytą rolę odgrywaną przez CT w dowolnym OPT, kiedy wspominamy CT, mamy zamiar wyznaczyć specjalną odpowiadającą OPT. Obecny pogląd, w którym świadomość jest podstawą informacji i fizyki opartej na informacji kwantowej, nazwiemy panpsychizmem informacji kwantowej. Zamiast QT można rozważyć postkwantową OPT–np. RQT (QT na rzeczywistych przestrzeniach Hilberta), FQT (fermionowy QT), PRB (OPT zbudowany na skrzynkach Popescu-Rohrlicha [26]) itp. [10], jednak niektóre cechy obecnej teorii świadomości można przełożyć na drugiej OPT, o ile pojęcie „splątania" przetrwa w rozważanej OPT.

Nieredukcyjne rozwiązanie psychoinformacyjne: zasady ogólne

Zasadniczy charakter zaproponowanego tutaj rozwiązania trudnego problemu został zasugerowany przez Davida Chalmersa jako spełniający następujące wymagania [1]:

Chalm1: Świadomość jako istota podstawowa, której nie da się wytłumaczyć w żaden prostszy sposób…

Chalm2: nieredukcyjna teoria doświadczenia określi podstawowe zasady, które mówią nam, w jaki sposób podstawowe doświadczenie zależy od fizycznych cech świata.

Chalm3: Te zasady psychofizyczne nie będą kolidować z prawami fizycznymi (zamknięcie fizyki). Będą raczej uzupełnieniem teorii fizycznej.

W ramach teorii informacji, w której fizyka opiera się na informacji, naszą zasadą będzie psychoinformacja (patrz Chalm3).

Nieredukcyjne

(Chalm2) Zaproponowana tutaj zasada psychoinformacyjna jest następująca:

P1: Zasada psychoinformacyjna: Świadomość jest doświadczeniem systemu informacyjnego dotyczącym jego własnego stanu informacyjnego i przetwarzania.

Jak wkrótce się przekonamy, niezwykle ważne jest, aby rodzaj informacji, której bezpośrednio doświadczamy, był kwantowy. Zasada P1 może wydawać się ad hoc, ale to samo dzieje się z wprowadzeniem dowolnej wielkości podstawowej (Chalm1,2) do fizyki, np. pojęć masy bezwładności, ładunku elektrycznego itp. Zasada P1 głosi, że doświadczenie jest podstawową cechą informacji , a więc i fizyki, która na nim superweniuje. P1 nie jest redukcyjny (Chalm2) i nie ma wpływu na fizykę (Chalm3), ponieważ rodzaj informacji związany z fizyką jest kwantowy + klasyczny. Z drugiej strony P1 uzupełnia fizykę (Chalm3), gdyż ta ostatnia superweniuje w teorii informacji. Naturalne jest teraz pytanie: jakie to systemy? Rzeczywiście, informacja jest wszędzie: światło pada na przedmioty, a następnie dociera do naszych oczu, dostarczając nam informacji o tych obiektach: kolorze, położeniu, kształcie itd.; informacja jest wspierana przez szereg systemów: tryby światła, następnie siatkówkę, następnie nerwy optyczne, a na koniec kilka oddolnych i odgórnych procesów wzrokowych zachodzących w mózgu. Chociaż ostateczna odpowiedź może wymagać eksperymentów z zakresu neuronauki, biologii molekularnej i nauk kognitywnych, możemy wykorzystać obecne podejście OPT do zainspirowania kluczowych eksperymentów. OPT ma tę moc, że jest metodą czarnej skrzynki, która nie wymaga szczegółowej „fizycznej" specyfikacji systemów, co jest wielką zaletą! Co więcej, nasza metoda rozwiązuje problem w kategoriach czysto merytorycznego rozumowania, niezależnego informacji pomocniczej „sprzętu", dokładnie tak, jak to robimy w teorii informacji. Taka niezależność sprzętu sprawia, że podejście to szczególnie dobrze nadaje się do rozwiązania problemu tak podstawowego, jak problem świadomości. Przejdźmy teraz do drugiej zasady:

P2: Zasada prywatności: doświadczeniem nie można się dzielić, nawet co do zasady.

Zasada P2 odgrywa szczególną rolę w wyborze teorii informacji zgodnych z teorią świadomości. Naszym doświadczeniem rzeczywiście nie można się dzielić: to fakt. Stawiamy hipotezę, że brak możliwości dzielenia się doświadczeniem obowiązuje w zasadzie, a nie tylko jako ograniczenie technologiczne. Kluczowym faktem jest to, że możliwość dzielenia się informacjami jest równoznaczna z czytelnością informacji bez zakłóceń. Niedawno udowodniono, że jedyną teorią, w której można wydobyć jakąkolwiek informację bez zakłóceń, jest klasyczna teoria informacji. Dochodzimy do wniosku, że P2 implikuje, że teoria świadomości potrzebuje nieklasycznej teorii informacji, a mianowicie QT lub „post-kwantowej" OPT. Tutaj rozważymy najbardziej znaną instancję OPT, a mianowicie QT. Jak się przekonamy, taki wybór teorii okazuje się bardzo skuteczny w wyjaśnianiu wszystkich głównych cech świadomości. Podajemy ten wybór teorii jako zasadę:

P2': Kwantowość doświadczenia: informacyjna teoria świadomości jest teorią kwantową.

Należy pamiętać, że teoria klasyczna zawsze towarzyszy każdemu OPT, więc będą miały miejsce wymiany i konwersje informacji klasycznej i kwantowej. Wprowadzimy teraz trzecią zasadę:

P3: Zasada psychoczystości: stan układu świadomego jest czysty

P3 może wydawać się arbitralne, jeśli błędnie utożsamiamy doświadczenie z wiedzą o nim. Rzeczywiste doświadczenie jest ontyczne i określone. Jest ontyczny, a la Kartezjusz (Cogito, ergo sum…). Istnienie

naszego doświadczenia jest z pewnością czymś, co do czego każdy się zgodzi, czymś, czego możemy być pewni. Jest to określone w tym sensie, że w danej chwili ma się tylko jedno doświadczenie, a nie rozkład prawdopodobieństwa doświadczeń. Ten ostatni opisywałby wiedzę wynikającą z cudzego doświadczenia. Doświadczenie zamazanej monety jest tak samo określone, jak doświadczenie ostrej monety, ponieważ są to dwa różne, określone doświadczenia. Określona jest znajomość lica monety, a można to osiągnąć tylko wtedy, gdy obraz monety jest wystarczająco ostry, w przeciwnym razie byłby on przedstawiony w kategoriach probabilistycznych, np. przez stan mieszany 1/2 ORZEŁ + 1/2 RESZKA, w oparciu o hipotezę uczciwej monety .

Możemy zatem zrozumieć, że zgodnie z tą samą definicją pojęcie mieszaniny jest pojęciem epistemicznym, opartym na naszej wcześniejszej wiedzy o monecie mającej dwa możliwe stany. W skrócie:

1. Doświadczenie opisuje ontyczny stan kwantowy, który jest czysty.

2. Wiedzę opisuje epistemiczny stan kwantowy, który jest na ogół mieszany.

Zobaczymy później, jak zasada P3 jest kluczowa dla rozwiązania problemu kombinacji. Podamy teraz naszą czwartą zasadę, niezależną od teorii:

P4: Zasada qualiów: doświadczenie składa się z ustrukturyzowanych qualiów.

Qualia (fenomenalne właściwości jakościowe) są elementami składowymi świadomego doświadczenia. Ich istnienie pozostawało jak dotąd niewytłumaczalne w ramach tradycyjnych ram naukowych. Ich prywatność i niewysłowiona sława są powszechnie znane. Qualia mają strukturę: niektóre z nich są bardziej podstawowe niż inne, np. kolory widma, wysokość dźwięku, pięć różnych rodzajów kubków smakowych (słodki, słony, kwaśny, gorzki i umami), doznania somatyczne (ból, przyjemność itp.).), emocji podstawowych (smutek, szczęście itp.). Qualia mają charakter kompozycyjny, a struktury wewnętrzne na ogół częściowo determinują ich charakter jakościowy. Łączą je odmienne relacje, tworząc liczne złożone struktury, takie jak myśli i emocje. Skoro świadomość postrzegamy jako bezpośredni owoc bardzo ustrukturyzowanego rodzaju informacji kwantowej (jesteśmy systemem podtrzymującym taką informację końcową), to zgodnie z zasadą P1 qualia odpowiadają stanom informacyjnym „odczuwanym” przez świadomy system, a zgodnie z zasadą P1 zasad P2 i P2’, stany takie są stanami kwantowymi takich układów.

S1: Qualia są opisywane przez ontyczne czyste stany kwantowe.

Zobaczymy w rozdz. 3 jak splątanie kwantowe może wyjaśniać organizację nietrywialnych jakości. Przestrzeń jakości byłaby zatem przestrzenią Hilberta systemów zaangażowanych w jakość. Klasyczna informacja docierająca ze środowiska za pośrednictwem zmysłów jest ostatecznie przekształcana w informację kwantową, która jest doświadczana jako qualia. Odwrotna transformacja kwantowa do klasycznej jest również kluczowa przy przekształcaniu struktur w ramach jakości w reprezentacje logiczne i geometryczne, które można wyrazić za pomocą klasycznych informacji, aby przekazać

klasycznym strukturom wybory wolnej woli. W tym celu niezbędna może być interakcja z pamięcią przeszłych doświadczeń, gdyż, jak zobaczymy, pamięć jest klasyczna. Wreszcie dochodzimy do problemu wolnej woli. Najpierw musimy sprecyzować, że wolna wola, w przeciwieństwie do świadomości, wytwarza skutki publiczne, które są klasycznymi przejawami informacji kwantowej. Ponieważ manifestacja jest klasyczna, możliwe wybory świadomego podmiotu są w zasadzie łącznie doskonale rozróżnialne. Następnie określamy, co rozumiemy pod pojęciem „bezpłatny".

S2: Wola jest wolna, jeśli jej nieprzewidywalność przez zewnętrznego obserwatora nie może być interpretowana w kategoriach braku wiedzy.

W szczególności ze stwierdzenia S2 wynika, że faktyczny wybór podmiotu korzystającego ze swojej wolnej woli nie może co do zasady być przewidywany z całą pewnością przez żadnego obserwatora zewnętrznego. Od razu widać, że jeśli uznać wybór za zmienną losową, to nie może on być klasyczny, co zawsze można zinterpretować jako brak wiedzy.Z drugiej strony idealnie pasuje do przypadku losowości kwantowej, której nie można tak interpretować.

Qualia: rola splątania

Zgodnie ze stwierdzeniem S1, qualia opisywane są ontycznymi stanami kwantowymi i będąc czystymi stanami, możemy je przedstawić za pomocą znormalizowanych wektorów $|\psi \in HA$ w przestrzeni Hilberta HA układu A. Dla ilustracji rozważymy dwa proste przypadki qualia: kierunek i kolor. Przykładów tych nie należy rozumieć dosłownie8, lecz jedynie w celach informacyjnych w celu zilustrowania koncepcji, że liniowe kombinacje jakości mogą dać początek nowym złożonym jakościom. Powiedzieliśmy, że qualia łączą się, tworząc nowe jakości, a same myśli i emocje są jakościami ustrukturyzowanymi. Nałożenie na siebie dwóch różnych rodzajów qualiów w splątany sposób tworzy nowe rodzaje qualiów. Rzeczywiście, rozważ nałożenie czerwonej strzałki w górę z zieloną strzałką w dół. Nie jest to żółta strzałka w prawo, jak można by się spodziewać, ponieważ ta ostatnia odpowiada niezależnej superpozycji kierunku i koloru, jak w następującym równaniu:

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |\bullet\rangle + |\downarrow\rangle \otimes |\bullet\rangle) \neq \frac{1}{2}(|\uparrow\rangle + |\downarrow\rangle) \otimes (|\bullet\rangle + |\bullet\rangle) = |\rightarrow\rangle \otimes |\bullet\rangle \tag{1}$$

Zamiast tego mielibyśmy

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |\bullet\rangle + |\downarrow\rangle \otimes |\bullet\rangle) = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) = |\bigstar\rangle, \tag{2}$$

gdzie ket z niebieską gwiazdą reprezentuje zupełnie nowe qualia. W bardziej ogólnym przypadku mamy wektor stanu z potrójnym, poczwórnym lub większym splątaniem we współczynniku iloczynu tensorowego, a bardziej ogólnie każdy układ jest splątany, np.

$$.\otimes(|a_1\rangle\otimes|b_1\rangle\otimes|c_1\rangle\otimes|d_1\rangle \pm |a_2\rangle\otimes|b_2\rangle\otimes|c_2\rangle\otimes|d_2\rangle)\otimes(|\uparrow\rangle\otimes|\bullet\rangle \pm |\downarrow\rangle\otimes|\bullet\rangle)\otimes\ldots \tag{3}$$

Widzimy, jak można w ten sposób uzyskać nowe rodzaje jakości, których liczba rośnie wykładniczo wraz z liczbą systemów. W rzeczywistości liczba różnych sposobów splątania N układów odpowiada liczbie podziałów N na liczby całkowite pomnożonej przez liczbę permutacji układów i dlatego rośnie jako N!Exp(√ 2N/3), a to bez uwzględnienia wektorów zmiennych, które można splątać! Ponieważ qualia

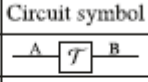
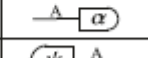
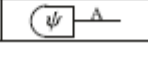
odpowiadają czystym stanom świadomych systemów, ich przestrzeń Hilberta pokrywa się z wielowymiarową przestrzenią Hilberta systemu. Eksperymentalnie, jeden być może będziemy w stanie zlokalizować systemy na podstawie wzorców neuronowych, ale nigdy nie będziemy w stanie odczytać zakodowanych informacji bez zniszczenia doświadczenia danej osoby, uzyskując w najlepszym przypadku tylko jedną uzupełniającą stronę jakości z wykładniczo wielu. W skrócie:

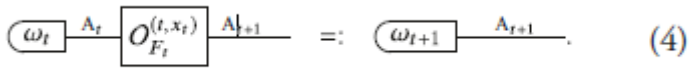
• Możliwa jest identyfikacja systemu, ale nie „doświadczenia" w nim zawartego.

Możliwość identyfikacji systemu informacyjnego dowodzi, że tożsamość obserwatora/agenta jest publiczna i tym samym jest skorelowana z jego „poczuciem siebie", które jest raczej prywatne. Jest to kluczowy wymóg jednolitej teorii świadomości oraz wolna wola, a mianowicie, że obserwatora/agenta można zidentyfikować zarówno prywatnie – od wewnątrz i poprzez qualia – jak i publicznie – bez informacji i poprzez informacje. W ramach QT (lub postkwantowych OPT) jest to możliwe.

Ewolucja świadomości i wolnej woli

Ponieważ stan kwantowy świadomego układu musi być zawsze czysty, jedynym sposobem zagwarantowania, że ewoluujący stan układu pozostanie czysty, jest czysta ewolucja (technicznie rzecz biorąc, jest ona atomowa, tj. jej mapa CP $T = \Sigma_i T_i \cdot T^{\dagger}_i$ ma pojedynczy operator Kraussa T_{i0}). Należy zauważyć, że zarówno stany, jak i efekty są również transformacjami odpowiednio z systemów trywialnych do nietrywialnych i odwrotnie. W tabeli 1 przedstawiamy teoretyczne reprezentacje trzech rodzajów transformacji bytowych, w tym szczególne przypadki stanu i skutku. Rozważmy teraz ogólny scenariusz złożonej świadomości system w stanie ontycznym ω_t w chwili t, ewoluujący poprzez jednoetapową transformację ontyczną $O^{(t,xt)}_{Ft}$ z wynikiem F_t, zależnym od klasycznego wejścia x_t ze zmysłów i pamięci

	Circuit symbol	Symbol	Map	Operator	Domain-codomain
transformation		$\mathcal{T}$	$\mathcal{T} = T \cdot T^{\dagger}$	T	$\mathrm{Bnd}(\mathcal{H}_A \to \mathcal{H}_B)$
Effect		$\langle\alpha\vert$	$\mathrm{Tr}[\cdot\vert\alpha\rangle\langle\alpha\vert]$	$\langle\alpha\vert$	$\mathrm{Bnd}(\mathcal{H}_A \to \mathbb{C})$
State		$\vert\psi)$	$\vert\psi\rangle\langle\psi\vert$	$\vert\psi\rangle$	$\mathrm{Bnd}(\mathbb{C} \to \mathcal{H}_A)$



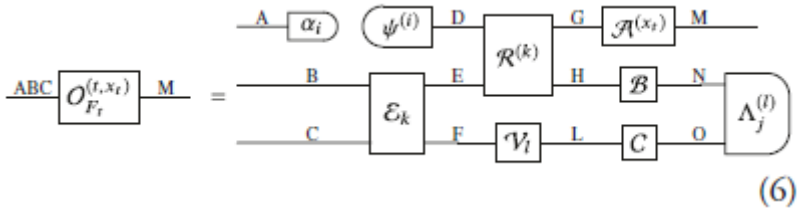
$$(4)$$

Transformacja epistemiczna byłaby sumą wszystkich bytowych odpowiadających wszystkim możliwym wynikom:

$$\mathcal{E}^{(t,x_t)} := \sum_{F_t} \mathcal{O}^{(t,x_t)}_{F_t}. \qquad (5)$$

Wynik F_t jest klasycznym wyjściem i utożsamiamy go z wolną wolą doświadczającego systemu. Jest to wynik probabilistyczny, który zależy od wcześniejszej historii jakości systemu. Jej przypadkowość ma charakter kwantowy, co oznacza, że nie można jej interpretować jako braku wiedzy i jako taka jest darmowa. Zauważ, że zarówno matematycznie, jak i dosłownie, wolna wola jest wynikiem transformacji, która odpowiada zmianie doświadczenia obserwatora/agenta. Konwersja informacji z kwantowej na klasyczną może uwzględniać także etap „wiedzy woli" odpowiadający „zamiarowi/celowi", czyli „rozumienie", jakie działanie jest podejmowane. Być może będziemy musieli

zapewnić bardziej wyrafinowaną reprezentację jednoetapowej transformacji ontycznej ewolucji w kategoriach obwodu kwantowego, na przykład:

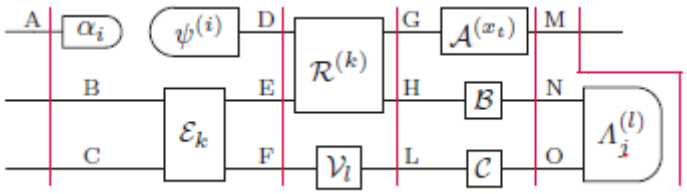


(6)

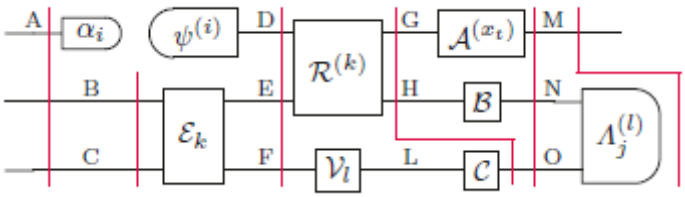
Generalnie dla każdego czasu t mamy inny obwód. W przykładzie w obwodzie (6) widzimy, że pojedynczy stopień może również zawierać stany i efekty, a układ wyjściowy całego obwodu (w tym przypadku M) jest generalnie inny (nawet nie jest izomorficzny) z układami wejściowymi ABC. Zgodnie z konwencją przy transformacji ontycznej $O^{(t,xt)}_{Ft}$ dolny indeks jest wynikiem losowym a górny indeks (t, xt) jest parametrem, na podstawie którego następuje transformacjageneralnie zależy. Ogółem w obwodzie mamy wolną wolę F_t = {i, k, l, j}, natomiast transformacje $\psi^{(i)}$,$R^{(k)}$,B i C są deterministyczne (nie mają dolnego indeksu), oraz nie przyczyniajcie się do wolnej woli; to samo dotyczy transformacji $A^{(xt)}$, która zależy od sygnału wejściowego x_t. W obwodzie (6) widać, że działa także informacja klasyczna, gdyż np. transformacja $R^{(k)}$ zależy od wyniku k transformacji E_k. Podobnie wybór testu $\{\Lambda^{(l}{}_{)j}\}$ zależy od wyniku l V_l, a stan $\psi^{(i)}$ zależy od wyniku efektu αi. Każdy element obwodu jest ontyczny, czyli atomowy, a atomowość składu sekwencyjnego i równoległego gwarantuje, że cała transformacja sama w sobie jest atomowa. Oznacza to, że jego operator Kraussa Ot można zapisać jako iloczyn operatorów Kraussa transformacji składowych w następujący sposób

$$O^{(t,x_t)}_{F_t} = \underbrace{\left(I_M \otimes \left\langle \Lambda^{(l)}_j \right|\right)}_{\mathrm{MNO}\to\mathrm{M}} \underbrace{\left(A^{(x_t)} \otimes B \otimes C\right)}_{GHL\to\mathrm{MNO}} \underbrace{\left(R^{(k)} \otimes V_l\right)}_{\mathrm{DEF}\to GHL} \underbrace{\left(|\psi_i\rangle\langle a_i| \otimes E_k\right)}_{ABC\to DEF}, \tag{7}$$

gdzie $O^{(t,xt)}_{Ft} = O^{(t,xt)}_{Ft} \cdot O^{(t,xt)\dagger}_{Ft}$. Zauważ, że w wyrażeniu dla operatora $O^{(t,xt)}_{Ft}$ w równaniu. (7) operatory od wejścia do wyjścia są zapisywane od prawej do lewej, tak jak tworzymy operatory na przestrzeniach Hilberta. Co więcej, wyrażenie (7) $O^{(t,xt)}_{Ft}$ nie jest jednoznaczne, gdyż zależy od wyboru foliacji obwodu, czyli od sposobu pokrycia wszystkich przewodów listkami w celu podzielenia obwodu na sekcje wejścia-wyjścia. Na przykład równanie. (7) odpowiadałoby foliowaniu na rysunku.



Inną foliację przedstawiono na przykład na rusinku, odpowiadającą wyrażeniu dla $O^{(t,xt)}_{Ft}$



$$O_{F_t}^{(t,x_t)} = \underbrace{\left(I_C \otimes \left\langle \Lambda_j^{(l)} \right|\right)}_{\text{MNO}\to\text{M}} \underbrace{\left(A^{(x_t)} \otimes B \otimes I_O\right)}_{\text{GHO}\to\text{MNO}} \underbrace{\left(R^{(k)} \otimes CV_l\right)}_{\text{DEF}\to\text{GHO}} \underbrace{\left(|\psi_i\rangle \otimes E_k\right)}_{\text{BC}\to\text{DEF}} \underbrace{\left(\langle a_i| \otimes I_{BC}\right)}_{\text{ABC}\to\text{BC}} \tag{8}$$

Przypomnijmy, że wszystkie operatory (w tym kets i bras jako operatory z i do trywialnej przestrzeni Hilberta C) są skurczami, tj. mają normę ograniczoną przez 1, odpowiadającą prawdopodobieństwom krańcowym nie większym niż 1. Zatem także $O^{(t,xt)}_{Ft}$ samo w sobie jest skurczem. Kontraktywność dla operatora X można wygodnie wyrazić jako

$$X \in \text{Bnd}(\mathcal{H}_{in} \to \mathcal{H}_{out}), \quad X^\dagger X \le P_{S_X} \le I_{in}, \quad P_{S_X} := \text{Proj Supp}\, X. \tag{9}$$

Teraz, ponieważ ewolucja świadomości musi być zawsze atomowa, możemy zapisać całą historię świadomości jako iloczyn operatorów Kraussa Ot ≡ $O^{(t,xt)}_{Ft}$ we wszystkich poprzednich momentach t

$$\Omega_t := O_t O_{t-1} \ldots O_1, \quad O_t := O_{F_t}^{(t,x_t)}, \tag{10}$$

i zastosuj operator historii do wektora falowego początkowego wektora stanu ontycznego $|\omega_0\rangle$

$$|\omega_t\rangle = \Omega_t |\omega_0\rangle. \tag{11}$$

Norma kwadratowa $||\omega t||^2$ wektora ω_t to prawdopodobieństwo pełnej historii stanów świadomych $\{\omega_0, \omega_1, \omega_2, \ldots, \omega_t\}$, a także historii wolnej woli $\{F_1, F_2, \ldots, F_t\}$

$$p(\omega_0, \omega_1, \omega_2, \ldots, \omega_t) = p(\emptyset, F_1, \ldots, F_t) = \|\omega_t\|^2 = \|\Omega_t \omega_0\|^2. \tag{12}$$

Pamięć

Ontyczna ewolucja stanu świadomości, choć zachowuje spójność, może zachować bardzo ograniczoną kwantową pamięć doświadczenia. Ta ostatnia, ze względu na kurczliwość O, będzie spadać bardzo szybko, gdy $||\Omega_t||^2 \simeq ||O||^{-2t}$, czyli będzie maleć wykładniczo wraz z liczbą kroków czasowych. Zamiast tego pamięć kwantowa nieodłącznie związana z ewolucją ontyczną działa dobrze jako krótkotrwały bufor umożliwiający zbudowanie pełniejszego doświadczenia – np. krajobrazu lub szczegółowego obiektu, a nawet wykrycie ruchu. Ile kubitów utworzy taki jednoetapowy bufor? Odpowiedź brzmi: niewiele.11 Rzeczywiście, jeśli otworzymy oczy na sekundę, aby popatrzeć na nieznaną scenę, a następnie zostaniemy poproszeni o udzielenie odpowiedzi na pytania binarne, otrzymamy tylko tuzin poprawnych odpowiedzi, lepszych od przypadku. Tor Nørretranders pisze [31]: Szerokość pasma świadomości jest znacznie mniejsza niż szerokość naszych zmysłów. … Świadomość składa się w znacznie większym stopniu z odrzuconych informacji niż z informacji obecnych. W naszej świadomości nie ma prawie żadnej informacji. Możemy rozsądnie wywnioskować, że w świadomości nie ma już miejsca na pamięć długoterminową i wywnioskować, że:

S3: Pamięć jest klasyczna. Jedynie krótkotrwały bufor umożliwiający zebranie każdego doświadczenia jest kwantowy.

Przez „krótki” rozumiemy czas porównywalny do czasu, w którym gromadzimy pełne doświadczenie, czyli rzędu sekundy.

Przeniesienie doświadczenia kwantowego do pamięci klasycznej

Jeśli mamy doświadczenia kwantowe i pamięć klasyczną, musimy przekształcić informację z kwantowej na klasyczną w procesie zapamiętywania oraz z klasycznej na kwantową w procesie przypominania. Pierwszy proces, czyli przeniesienie doświadczenia kwantowego do pamięci klasycznej, z konieczności musi być niekompletny, w przeciwnym razie naruszyłby twierdzenie o kwantowym nieklonowaniu. Twierdzenie Holevo [33] stwierdza, że maksymalna ilość klasycznej informacji, jaką można wydobyć z układu kwantowego, to liczba bitów równa liczbie kubitów tworzących układ kwantowy. Oczywiście taka maksymalna klasyczna informacja jest nieskończenie mała w porównaniu z kontinuum klasycznej informacji potrzebnej do klasycznego przekazania stanu kwantowego! Z pojedynczego pomiaru można wydobyć klasyczną informację tylko o jednym z kontinuum uzupełniających się aspektów stanu kwantowego, np.: wzdłuż danego kierunku dla obrotu. Co więcej, stan po pomiarze musiałby zostać zakłócony13 ze względu na kompromis między informacją a zakłóceniem.14 Jaki zatem pomiar należy wykonać, aby zebrać najlepszą klasyczną informację z układu kwantowego? Odpowiedź jest łatwa dla pojedynczego kubitu. Po prostu wykonaj zwykły pomiar von Neumanna w losowym kierunku! Zostało to udowodnione w [35].15 Dla wyższych wymiarów w zasadzie możliwe jest uogólnienie tego wyniku. Przestrzeń prawdopodobieństwa okazuje się jednak geometrycznie znacznie bardziej skomplikowana niż kula16, a schemat losowego wyboru jest nieznany. Z tego powodu większą klasę testów obserwacyjnych można uznać za kompletne informacyjnie („infokompletne”)17, które w odróżnieniu od przypadku optymalnego pomiaru kierunku wirowania można uznać za dyskretne i o skończonej liczbie wyników. Taki pomiar odpowiadałby następującej transformacji epistemicznej:

$$\mathcal{M} \in \mathrm{Tm}_1(\mathrm{A} \to \mathrm{B}), \quad \mathcal{M} = \sum_{j \in J} M_j \rho M_j^\dagger, \quad |J| \geq \dim(\mathcal{H}_\mathrm{A})^2, \tag{13}$$

co jest gruboziarnistym niepełnym testem kwantowym $\{M_j\}_{j\in J}$, gdzie $M_j = M_j \cdot M^{\dagger}_j$, gdzie $\{|M|^2_j\}_{j\in J}$ jest niepełny test obserwacji. „Infocomplete” oznacza, że w granicach nieskończenie wielu zastosowań w tym samym regenerowanym stanie, pomiar pozwala na wykonanie pełnej tomografii stanu. Pomiar infokompletny (lub obserwowane wcześniej losowe obserwacje) lepiej nadaje się do wydobywania klasycznych informacji z bufora kwantowego pamięci długotrwałej, ponieważ nie uprzywilejowuje żadnej konkretnej obserwowalnej. Jest również prawdopodobne, że świadomy akt zapamiętywania doświadczenia można osiągnąć poprzez faktyczne wielokrotne przygotowanie stanu bytowego w buforze kwantowym i wielokrotne wykonanie testu niepełnego, dając w ten sposób możliwość zapamiętania (ewentualnie częściowej) tomografii państwo. Zauważ, że ogólnie stan ontyczny $M_j\rho M^{\dagger}_j$ dla danego j nadal zależy od ρ. Jednak test go zakłóci, a im mniej zakłóceń, tym mniejszą ilość klasycznej informacji można wydobyć.

Przeniesienie pamięci klasycznej na doświadczenie kwantowe

Widzieliśmy, że do (w przybliżeniu) przechowywania doświadczenia w klasycznej pamięci długotrwałej potrzebny jest pomiar infokompletny. Z definicji odzyskanie doświadczenia wymaga jego odtworzenia, co oznacza, że odpowiedni stan bytowy jest (w przybliżeniu) odtwarzany na podstawie klasycznych przechowywanych danych. Wspomnienie, jako klasyczne, zostanie odczytane bez zakłóceń, dzięki czemu będzie dostępne do dalszych wspomnień. W celu przeniesienia informacji klasycznej na kwantową zaproponowano metody przygotowania stanu w oparciu o schematy obwodów kwantowych. Możliwym punktem odniesienia dla procesu przechowywania i przywoływania pamięci

jest maksymalna wierność, jaką można w zasadzie osiągnąć za pomocą schematu pomiaru i ponownego przygotowania, który optymalizuje wierność między stanem doświadczonym a stanem przywołanym. Rozważymy M ≥ 1 kopii wejściowych tego samego stanu, aby określić ilościowo uwagę na doświadczenie. Optymalną wierność między stanem doświadczonym a stanem przywołanym, uśrednioną po wszystkich możliwych doświadczeniach (tj. Stanach wejściowych), określa wzór

$$F(A, M, d) = \frac{M+1}{M+d}, \qquad (15)$$

gdzie d jest wymiarem systemu doświadczającego. Taka górna granica wierności może zostać wykorzystana do wywnioskowania efektywnego wymiaru d systemu związanego ze świadomością, np. w wysoce skupionych, ograniczonych doświadczeniach, takich jak te związane z maskowaniem świadomej percepcji.

Transfer informacji z ciała do świadomości i odwrotnie

Oczekiwalibyśmy, że większość działań ludzkiego ciała będzie automatyczna i będzie wykorzystywać klasyczne informacje, które nigdy nie są przekładane na informacje kwantowe, których świadomość może doświadczyć. Część klasycznej informacji wytwarzanej przez ciało, która wymaga przekształcenia w informację kwantową, powinna stanowić jedynie najistotniejszą informację, która wspiera percepcję i zrozumienie jakości niezbędnych do „życia" i dokonywania odpowiednich wyborów z wolnej woli. Ilość informacji kwantowej, którą należy przełożyć na język klasyczny dla celów wolnej woli, kontrola ciała od góry do dołu powinna być stosunkowo niewielka.

O problemie kombinacji

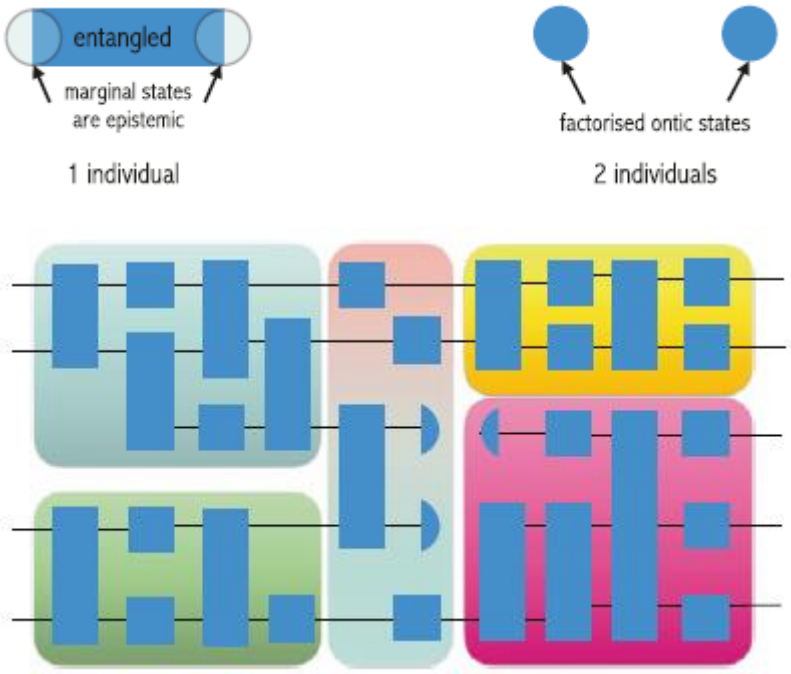
Problem kombinacji dotyczy kwestii, czy i w jaki sposób podstawowe świadome umysły zaczynają tworzyć, konstytuować lub powodować powstanie innego, dodatkowego świadomego umysłu. Z definicji problem staje się kluczowy dla panpsychizmu: jeśli świadomość jest wszędzie, jakie jest kryterium wyboru nowych świadomych jednostek? Czy zjednoczenie dwóch świadomych istot jest bytem świadomym? Jeśli to prawda, wówczas każdy podzbiór istoty świadomej może być również istotą świadomą. Obecne podejście teoretyczne dostarcza precyzyjnego kryterium indywiduacji. Kryterium to wywodzi się z zasady P3 o czystości kwantowych stanów świadomych, a co za tym idzie, o konieczności transformacji ontycznych.

Zobaczmy jak to działa.

Rozsądne jest stwierdzenie, że jednostkę definiuje ciągłość jej doświadczeń. Dla niektórych czytelników takie stwierdzenie może być od razu oczywiste. Jednakże dla tych, którzy mogą się nie zgodzić, proponujemy eksperyment myślowy. Rozważmy futurystyczny „eksperyment z teleportacją kwantową", co oznacza, że stan kwantowy układu zostaje zastąpiony przez odległy układ izomorficzny. Materia (elektrony, protony itp.), z której składa się teleportowana osoba, jest dostępna w miejscu odbioru i w zasadzie jest nie do odróżnienia od tej samej materii w miejscu transmisji: w rzeczywistości teleportowany jest tylko jej stan kwantowy. Powstała osoba byłaby taka sama jak pierwotna, łącznie z jej myślami i pamięcią. Powyższy eksperyment myślowy sugeruje następujące kryterium indywiduacji w ramach naszego podejścia teoretycznego:

S4: Świadomy umysł jest złożonym systemem w stanie bytowym podlegającym transformacji bytowej, bez podsystemu jako takiego.

Na rysunkach ilustrujemy zastosowanie kryterium w dwóch paradygmatycznych przypadkach.



W tym miejscu podkreślamy, że dla kryterium kluczowa jest rola interakcji. W tej części opisano, co jest konieczne do utworzenia kombinacji świadomych bytów, usuwając w ten sposób trudności napotykane w modelach panpsychistycznych opartych na fizyce klasycznej. Kluczową ideą jest to, że stany łączących się układów i ich przekształcenia są ontyczne i że takie układy oddziałują kwantowo. Wymagania te zapewniają, że połączona istota również znajduje się w stanie czystym i żaden z jej podzbiorów nie jest świadomy. Jest oczywiste, że interakcje pomiędzy świadomą istotą a otoczeniem (w tym innymi bytami) mogą mieć jedynie charakter klasyczny, w przeciwnym razie powstałaby większa całość, która utrzymywałaby się tak długo, jak długo nie nastąpi rozplątująca transformacja, jak w przypadku zilustrowanym na rysunku.

Eksperymentowalność i symulowalność

Proponowanie wykonalnych eksperymentów dotyczących kwantowej natury świadomości jest bardzo ekscytującym wyzwaniem. W zasadzie kwantowość można ustalić poprzez eksperymentalną demonstrację jednej z dwóch głównych nieklasycznych cech teorii kwantów: nielokalności i komplementarności. Obydwa pojęcia nie są niezależne, ponieważ aby udowodnić nielokalność, oprócz wspólnego splątania potrzebujemy obserwacji uzupełniających. Aby udowodnić nielokalność świadomości, potrzebujemy pomiarów w dwóch oddzielnych punktach, wystarczająco odległych od siebie, aby wykluczyć związek przyczynowy, a taki wymóg jest bardzo trudny, ponieważ wymaga bardzo szybkich pomiarów, biorąc pod uwagę, że dla odległości 3 cm między pomiarami punktów, do sygnalizacji wystarczy różnica czasu wynosząca jedną dziesiątą nanosekundy. Jeśli chodzi o nielokalność w połączeniu z komplementarnością, spekulujemy, że mogą one być razem zaangażowane w wizję 3D i korzystamy z okazji, aby zasugerować, że taki kierunek badań może w każdym razie być owocnym polem eksperymentalnych badań nad świadomością. Na przykład autentyczny przypadek komplementarności w doświadczeniu 3D następuje w wyniku patrzenia na obrazy Magic Eye opublikowane w serii książek. Obrazy te zawierają autostereogramy, które pozwalają większości ludzi oglądać obrazy 3D poprzez skupianie n wzorów 2D, które wydają się nie mieć nic wspólnego z wyłaniającym się z nich obrazem 3D. Widz musi rozbiegać lub zbiegać swoje oczy, aby zobaczyć ukryty trójwymiarowy obraz we wzorach. Obraz 3D, który pojawia się w tym doświadczeniu, jest jak szklisty

obiekt, który w zależności od zbieżności lub rozbieżności oczu ukazuje się jako wklęsły lub wypukły. Jest oczywiste, że dwa alternatywne widoki 3D – wypukły i wklęsły – są naprawdę uzupełniającymi się doświadczeniami, a każde doświadczenie ma wewnętrzną całość. Jeśli chodzi o samą komplementarność, w literaturze rozważa się spekulatywne powiązania z kontrastującymi lub przeciwnymi wymiarami ludzkiego doświadczenia, m.in. „analiza" kontra „synteza" oraz „logika" kontra „intuicja". Tutaj jednak interesuje nas komplementarność doświadczeń możliwie powtarzalnych u różnych jednostek, typu gedanken-eksperymenty a la Heisenberg, np. jedno doświadczenie nie da się pogodzić z drugim i/lub przeszkadzają sobie nawzajem. Posiadanie czegoś więcej niż dwóch komplementarnych obserwabli, a mianowicie informacyjnie kompletnego zestawu pomiarów, takich jak wszystkie trzy macierze Pauliego dla pojedynczego kubitu świadomości, pozwoliłoby nam przeprowadzić tomografię kwantową [36] stanu kubitu, zakładając, że jest on ponownie przygotowany wiele razy, np. poprzez intensywną, długotrwałą uwagę. Według naszej wiedzy ocena wykonalności tego rodzaju eksperymentu jest podobna do trudności eksperymentu nielokalnego. Wreszcie, poza udowodnieniem kwantowości doświadczenia, możemy przynajmniej eksperymentalnie wywnioskować pewne parametry teoretyczne w podejściu teorii kwantowej zgodne z obserwacjami. Dzieje się tak m.in. eksperymentu wspomnianego już przy omawianiu górnej granicy (15) wierności przywoływania pamięci, którą można wykorzystać do wnioskowania o efektywnym wymiarze d systemu zaangażowanego w świadomość w wysoce skupionych, ograniczonych doświadczeniach, takich jak maskowanie świadomej percepcji]. Wierzymy, że eksperymenty z przywoływaniem pamięci w porównaniu ze zmiennymi, takimi jak uwaga, opóźnienie odzyskiwania pamięci i zmienne typy qualiów, mogą być pomocne we wstępnym mapowaniu wymiarowości przestrzeni zaangażowanych systemów świadomych. Na zakończenie przedstawimy kilka rozważań na temat wykonalności symulacji świadomego procesu, takiego jak zaproponowane tutaj. Jak już wspomniano, ze względu na czystość procesu ontycznego, symulacja wymaga jedynie mnożenia macierzy na ogół prostokątnych, co dla macierzy o tym samym wymiarze d jest zasadniczo procesem (d3). W przypadku rzadkich macierzy (co często ma miejsce w symulacji kwantowej) proces można znacznie przyspieszyć. Aby jednak wyznaczyć rozkład prawdopodobieństwa każdego kroku bytowego, należy oszacować prawdopodobieństwo warunkowe stanu wyjściowego poprzedniego kroku, a to wiąże się z pomnożeniem przez wszystkie możliwe wyniki (wolną wolę) odpowiedniego operatora Krausa ostatniego transformacja ontyczna. Przy pamięci RAM rzędu GB z pewnością można by operować na kilkunastu kubitach jednocześnie. Należy to porównać z 53 kubitami komputera kwantowego Google lub IBM, największego obecnie dostępnego, który prawdopodobnie udostępnia klasyczne informacje w tandemie z dużym klasycznym komputerem. Nie jest jednak wykluczone, że pewne szczególne zjawiska można było już odkryć/przeanalizować za pomocą laptopa.

Implikacje filozoficzne

Zaprezentowaliśmy teorię świadomości opartą na zasadach, założeniach i kluczowych koncepcjach, które uważamy za kluczowe dla solidności tej teorii i usunięcia ograniczeń większości teorii panpsychistycznych. Wierzymy, że nadanie wewnętrznej rzeczywistości i sprawczości systemom kwantowym w czystych stanach kwantowych, wraz z konwersją informacji z klasycznej na kwantową i odwrotnie, jest bezprecedensowe i ma poważne konsekwencje filozoficzne i naukowe. W obecnym podejściu wolna wola i świadomość idą ręka w rękę, umożliwiając systemowi działanie w oparciu o jego doświadczenie qualia poprzez konwersję informacji kwantowej na klasyczną, a tym samym nadanie mocy przyczynowej subiektywności – coś, co do tej pory było uważane za wysoce kontrowersyjne, jeśli nie niemożliwe. Teoria przewiduje, że świadomy podmiot może celowo przekształcić informację kwantową w konkretną informację klasyczną, aby wyrazić swoją wolną wolę, co jest klasycznym wyjściem, który jest w zasadzie nieprzewidywalny ze względu na swoje kwantowe pochodzenie. Teoria byłaby niespójna bez identyfikacji świadomego systemu pod względem czystości

i nierozłączności stanu kwantowego, który utożsamiany jest z doświadczeniem systemów. Czystość niedeterministycznej ewolucji kwantowej utożsamia świadomość z sprawczością poprzez jej wynik. Z metafizycznego punktu widzenia proponowaną interpretację, zgodnie z którą czysty, nierozdzielny stan kwantowy jest stanem świadomości, można wywrócić do góry nogami, przyjmując ontologię świadomości i sprawczości jako pierwotną, w wyniku czego fizyka wyłoniłaby się ze świadomości i sprawczości. Ta sama interpretacja uznałaby wówczas fizykę klasyczną za pełne urzeczowienie (uprzedmiotowienie) rzeczywistości kwantowej jako sprawstwo kwantowo-klasyczne odpowiadające wolnej woli świadomych istot istniejących całkowicie w sferze kwantowej. Ontologia wyprowadzona z przyjęcia świadomości jako fundamentalnej byłaby taka, że obiektywność i fizyka klasyczna superweniują w fizyce kwantowej, fizyka kwantowa superweniuje w oparciu o informację kwantową, a informacja kwantowa superweniuje w świadomości. Gdybyśmy zaakceptowali ten spekulacyjny pogląd, fizykę można by wówczas rozumieć jako opisującą nieograniczoną przyszłość, która jeszcze nie istnieje, ponieważ świadome podmioty nie dokonały jeszcze wyborów opartych na wolnej woli. W tej perspektywie my, jako istoty świadome, jesteśmy współtwórcami naszego świata fizycznego. Robimy to indywidualnie i zbiorowo, chwila po chwili i nie zdając sobie z tego sprawy, poprzez nasze wybory wolnej woli.

Załączniki dotyczące ogólnych opcji OPT

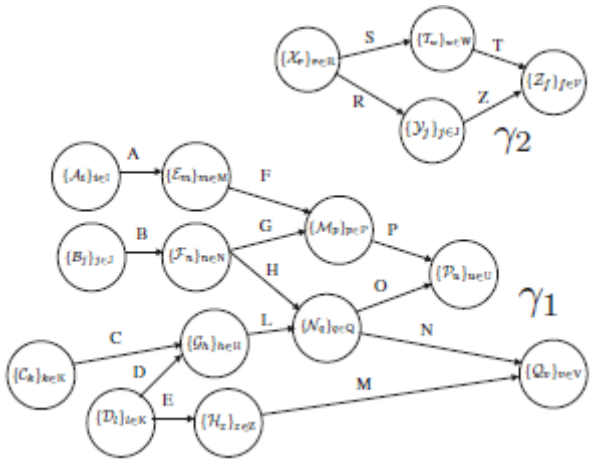
W tych dodatkach przedstawiamy ogólne operacyjne probabilistyczne ramy dla przyszłych postkwantowych badań teorii świadomości. Podajemy tutaj krótką ilustrację tego, co stanowi operacyjną teorię probabilistyczną (OPT), przy czym teoria kwantowa i teoria klasyczna są najbardziej odpowiednimi przykładami. Jak czytelnik może docenić, OPT zapewnia ramy, które są znacznie bogatsze, bardziej ogólne, elastyczne i bardziej rygorystyczne matematycznie niż inne ramy teoretyczne, takie jak podejście przyczynowe Tononiego

Ramy operacyjnej teorii probabilistycznej (OPT).

Nie będzie przesadą stwierdzenie, że ramy OPT reprezentują nową rewolucję galilejską dla nauki Tak naprawdę jest to pierwszy przypadek niezależny od teorii. Ustalono zbiór zasad dotyczących budowania teorii w fizyce i ewentualnie w innych naukach. Zasady takie stanowią tzw. ramy operacyjne. Jego rygor wynika z prostego faktu, że OPT to po prostu „metatematyka", ponieważ jest to rozdział teorii kategorii. Ściślej mówiąc, największa klasa OPT odpowiada kategorii o oplocie monooidalnym. Fakt, że zastosowano tę samą strukturę kategoryzacyjną w informatyce daje wyobrażenie o dokładności i zakresie stosowalności reguł OPT. W kilku nowatorskich artykułach Lucien Hardy przedstawił ramy heurystyczne, które można uznać za prekursora OPT, które po raz pierwszy pojawiły się i były wkrótce powiązane z podejściem kategorycznym w informatyce.Jak już wspomniano, zarówno QT, jak i CT są OPT, ale można zbudować inne OPT, takie jak odmiany QT, np. fermionowy QT lub QT w rzeczywistej przestrzeni Hilberta lub QT z tylko kubity, ale także CT ze splątaniem lub bez lokalnej rozróżnialności. Uważa się, że inne teorie zabawek, takie jak PR-Boxy, można uzupełnić OPT. Połączenie OPT z informatyką odzwierciedla ducha OPT, który zasadniczo był zrodził się w oparciu o nową dziedzinę informacji kwantowej. Rzeczywiście, ramy OPT to sformalizowanie zasad budowania obwodów kwantowych i kojarzenia z nimi wspólnego prawdopodobieństwa: w ten sposób OPT dosłownie staje się przedłużeniem teorii prawdopodobieństwa.

Ogólny pomysł

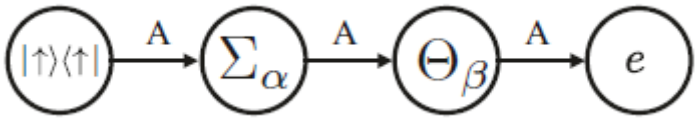
Jak działa OPT? Przypisuje każdemu wspólnemu prawdopodobieństwu wielu zdarzeń zamknięty skierowany graf acykliczny (CDAG) relacji wejście-wyjście, jak na rysunku



Każde zdarzenie (np. Em na rysunku) jest elementem pełnego testu ($\{Em\}_{m\in M}$ na rysunku) ze znormalizowanym prawdopodobieństwem krańcowym $\Sigma_{m\in M}p(E_m) = 1$. Wykres pokazuje, że krańcowy rozkład prawdopodobieństwa dowolnego zestawu testów w dalszym ciągu zależy od marginalizowanego zbioru, np. krańcowy rozkład prawdopodobieństwa test $\{E_m\}_{m\in M}$ zależy od pełnego wykresu testów, z którym jest powiązany, chociaż został częściowo lub całkowicie zmarginalizowany. Z reguły rozłączone wykresy (jak γ_1 i γ_2 na rysunku powyżej) są statystycznie niezależne, tj. ich rozkłady prawdopodobieństwa rozkładają się na czynniki. Zorientowane przewody oznaczające połączenia wyjściowo-wejściowe pomiędzy testami (oznaczone literami rzymskimi na rysunku powyżej) to tzw. układy teoretyczne.

Paradygmatyczny przykład kwantowy

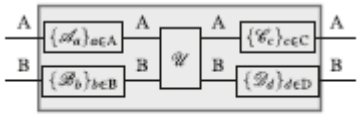
Paradygmatyczny przykład wykresu OPT pokazano na rysunku .

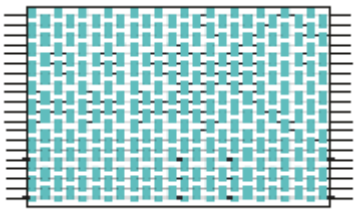


Mamy tam źródło cząstek, wszystkie o spinie (czyli w stanie $\rho = |\uparrow\uparrow|$). Na wyjściu mamy dwa pomiary von Neumanna w kaskadzie, pierwszy Σ_α mierzący σ_α, a drugi Θ_β mierzący σ_β, gdzie α i β mogą przyjmować dowolną z dwóch wartości x, z. Układ jest przedstawiony za pomocą wykresu pokazanego na rysunku, gdzie A oznacza układ odpowiadający spinowi cząstki, e test deterministyczny, który po prostu odrzuca cząstkę, oraz dwa testy Σ_α, Θ_β ($\alpha, \beta = x, z$) dwa pomiary von Neumana. Teraz wyraźnie dla $\alpha = z$ mamy krańcowy rozkład prawdopodobieństwa $p(\Sigma_z) = (1,0)$ i $p(\Sigma_x) = (1/2, 1/2$ niezależnie od wyboru testu Θ_β. Natomiast dla drugiego testu mamy prawdopodobieństwo krańcowe $p(\Theta_x) = (1,0)$ dla Σ_z i $p(\Theta_x) = (1/2 , 1/2)$ dla Σ_x. Dochodzimy do wniosku, że prawdopodobieństwo krańcowe Σ_α jest niezależne od wyboru testu Θ_β, natomiast prawdopodobieństwo krańcowe Θ_β zależy od wyboru Σ_α. Zatem prawdopodobieństwo krańcowe Θ_β generalnie zależy od wyboru Σ_α, a koncepcja ta wykracza poza treść wspólnego prawdopodobieństwa i wymaga wykresu OPT. Teoretycznie stwierdzamy, że z testu Σ_α do testu Θ_β „coś leci” (chociaż tego nie widzimy!): to właśnie teoretycznie opisujemy jako wirującą cząstkę! To dobrze ilustruje pojęcie układu: teoretyczne połączenie pomiędzy testowanymi zdarzeniami.

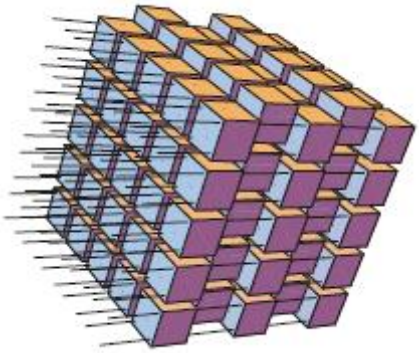
Podejście czarnoskrzynkowe

Wreszcie OPT jest podejściem czarnoskrzynkowym, gdzie każdy test jest opisany obiektem matematycznym, który można „faktycznie osiągnąć” za pomocą bardzo specyficznego urządzenia fizycznego. Nic jednak nie stoi na przeszkodzie, aby zapewnić bardziej szczegółową realizację testu OPT, np. jak na rysunku





Zauważ, że chociaż dowolny wykres można przedstawić w 2D (np. wykorzystując skrzyżowanie przewodów), to lepiej jest zaprojektować go w 3D (2D + wejście-wyjście), jak na rysunku.



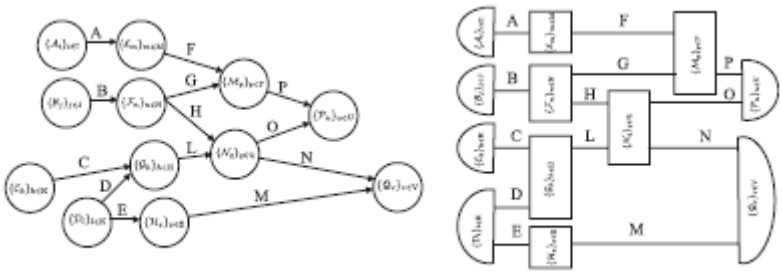
OPT i cel nauki

Szybko można się zorientować, że ramy OPT pozwalają precyzyjnie wyrazić najogólniejszy cel nauki, jakim jest łączenie zachodzących obiektywnych faktów (zdarzeń), stworzenie teorii takich „powiązań” (systemów), a tym samym umożliwienie przewidywania przyszłych zdarzeń pod względem łącznych prawdopodobieństw zdarzeń w zależności od ich powiązań. Jedną z głównych metodologicznie istotnych cech OPT jest to, że doskonale odróżnia „obiektywne dane” od „elementu teoretycznego”. Obiektywne są przeprowadzane testy i wynik każdego testu. Teorią jest wykres powiązań między testami, wraz z matematycznymi reprezentacjami obu systemów i testów. Ramy OPT dyktują zasady, które powinien spełniać opis matematyczny, a konkretny OPT podaje konkretny matematyczny reprezentacja systemów i testów/zdarzeń oraz ich składy (w kolejności i równolegle) w celu zbudowania CDAG.

OPT i metoda naukowa

Jedną z głównych zasad metody naukowej jest wyraźne rozróżnienie między tym, co eksperymentalne, a tym, co teoretyczne. Choć wydawałoby się to trywialnym stwierdzeniem, takie zamieszanie często staje się źródłem nieporozumień między naukowcami. Chociaż opis aparatu jest na ogół przeplatany pojęciami teoretycznymi, czyste dane eksperymentalne muszą mieć konwencjonalnie zdefiniowany „status obiektywności”, odpowiadający informacjom „powszechnie znanym”, a mianowicie możliwym do współdzielenia przez dowolną liczbę różnych obserwatorów. Wówczas zarówno język teoretyczny, jak i ramy muszą odzwierciedlać rozróżnienie teoria–eksperyment, wskazując wprost, którym pojęciom przypisuje się status obiektywności. Logika wraz z rachunkiem zdarzeń Boole'a jest istotną częścią języka, a teorię prawdopodobieństwa można uznać za rozszerzenie logiki, przypisując

prawdopodobieństwa zdarzeniom. Pojęciem promowanym do statusu obiektywności jest „wynik testu", ogłaszający, które zdarzenie w danym teście miało miejsce. Ramy OPT stanowią zatem rozszerzenie teorii prawdopodobieństwa, zapewniając teoretyczną łączność między zdarzeniami, inne teoretyczne składnikami są matematyczne opisy systemów i testów. OPT jako ogólna teoria informacji. Od razu widać, że CDAG jest dokładnie tym samym wykresem obwodu kwantowego, jaki rysuje się w informatyce kwantowej. Obwód kwantowy z kolei można interpretowany jako diagram przebiegu programu, w którym każdy test reprezentuje podprogram, a przewody reprezentują rejestry, poprzez które podprogramy przekazują dane. Rzeczywiście, OPT można ogólnie uważać za właściwą strukturę dla informatyki.



Aby zapoznać się z najnowszą, pełną prezentacją ram OPT, czytelnika odsyłamy do pracy lub nowszej i dogłębnej prezentacji

Notacja i skróty

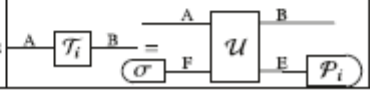
$\mathrm{Bnd}^+(\mathcal{H})$	bounded positive operators over $\mathcal{H}$
$\mathrm{CP}_{\leqslant}$	trace-non increasing completely positive map
$\mathrm{CP}_{=}$	trace-preserving completely positive map
$\mathcal{H}$	Hilbert space over $\mathbb{C}$
Cone(S)	conic hull of S
$\mathrm{Cone}_{\leqslant 1}(\mathrm{S})$	convex hull of $\{\mathrm{S} \cup 0\}$
Conv(S)	convex hull of S
Eff(A)	set of effects of system A
$\mathrm{Eff}_1(\mathrm{A})$	set of deterministic effects of system A
$\mathrm{Mrkv}_{\leqslant}$	normalization-non-increasing right-stochastic Markov matrices
Mrkv_1	normalization-preserving right-stochastic Markov matrices
$\mathrm{Prm}(n)$	$n \times n$ permutation matrices
$(\mathbb{R}^n)^+_{\leqslant 1}$	$\{\mathbf{x} \in \mathbb{R}^{n_\mathrm{A}} \mid \mathbf{x} \geq 0, \mathbf{x} \leqslant \mathbf{1}\}$: simplex $\mathbf{S}^{n_\mathrm{A}+1}$
$(\mathbb{R}^n)^+_1$	$\{\mathbf{x} \in \mathbb{R}^{n_\mathrm{A}} \mid \mathbf{x} \geq 0, \|\mathbf{x}\|_1 = 1\}$: simplex $\mathbf{S}^{n_\mathrm{A}}$
St(A)	set of states of system A
$\mathrm{St}_1(\mathrm{A})$	set of deterministic states of system A
$\mathrm{T}(\mathcal{H})$	trace-class operators over $\mathcal{H}$
$\mathrm{T}^+(\mathcal{H})$	trace-class positive operators over $\mathcal{H}$
$\mathrm{T}^+_{\leqslant 1}(\mathcal{H})$	positive sub-unit-trace operators over $\mathcal{H}$
$\mathrm{T}^+_{=1}(\mathcal{H})$	positive unit-trace operators over $\mathcal{H}$
$\mathrm{Trn}(\mathrm{A} \to \mathrm{B})$	set of transformations from system A to system B
$\mathrm{Trn}_1(\mathrm{A} \to \mathrm{B})$	set of deterministic transformations from system A to system B
$\mathbb{U}(\mathcal{H})$	unitary group over $\mathcal{H}$
	Special-case corollaries
	$\mathrm{T}(\mathbb{C}) = \mathbb{C}, \quad \mathrm{T}^+(\mathbb{C}) = \mathbb{R}^+, \quad \mathrm{T}^+_{\leqslant 1}(\mathbb{C}) = [0, 1], \quad \mathrm{T}^+_{=1}(\mathbb{C}) = \{1\}$ $\mathrm{CP}(\mathrm{T}(\mathcal{H}) \to \mathrm{T}(\mathbb{C})) = \mathrm{P}(\mathrm{T}(\mathcal{H}) \to \mathrm{T}(\mathbb{C})) = \{\mathrm{Tr}[\cdot E], E \in \mathrm{Bnd}^+(\mathcal{H})\}$ $\mathrm{CP}(\mathrm{T}(\mathbb{C}) \to \mathrm{T}(\mathcal{H})) = \mathrm{P}(\mathrm{T}(\mathbb{C}) \to \mathrm{T}(\mathcal{H})) = \mathrm{T}^+(\mathcal{H})$ $\mathrm{CP}_{\leqslant}(\mathrm{T}(\mathbb{C}) \to \mathrm{T}(\mathcal{H})) \equiv \mathrm{T}^+_{\leqslant 1}(\mathcal{H})$ $\mathrm{CP}_{\leqslant}(\mathrm{T}(\mathcal{H}) \to \mathrm{T}(\mathbb{C})) \equiv \{\epsilon(\cdot) = \mathrm{Tr}[\cdot E], 0 \leqslant E \leqslant I\}$ $\mathrm{Mrkv}_{\leqslant}(n, 1) = (\mathbb{R}^n)^+_{\leqslant 1}$ $\mathrm{Mrkv}_1(n, 1) = (\mathbb{R}^n)^+_{=1}$
	Abbreviations
CT	Classical theory
OPT	Operational probabilistic theory
PR boxes	Popescu-Rohrlich boxes
QT	Quantum theory

Teoria kwantowa

Minimalna matematyczna aksjomatyzacja teorii kwantowej jako OPT podano w Tabeli .

Quantum theory		
system	A	$\mathcal{H}_{\mathrm{A}}$
system composition	AB	$\mathcal{H}_{\mathrm{AB}} = \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$
transformation	$\mathcal{T} \in \mathrm{Trn}(\mathrm{A} \to \mathrm{B})$	$\mathcal{T} \in \mathrm{CP}_{\leqslant}(\mathrm{T}(\mathcal{H}_{\mathrm{A}}) \to \mathrm{T}(\mathcal{H}_{\mathrm{B}}))$
Born rule	$p(\mathcal{T}) = \mathrm{Tr}\,\mathcal{T}$	$\mathcal{T} \in \mathrm{Trn}(\mathrm{I} \to \mathrm{A})$

W przypadku OPT musimy podać matematyczny opis systemów, ich składu i transformacji z jednego systemu do drugiego. Następnie wszystkie zasady kompozycji transformacji i odpowiadających im systemów są dostarczane przez framework OPT. Czytelnik nieobeznany z takim frameworkiem może po prostu skorzystać z intuicyjnej konstrukcji obwodów kwantowych. W tabeli przedstawiamy następujące główne twierdzenia z aksjomatów. Odsyłamy czytelnika zainteresowanego motywacjami obecnej aksjomatyzacji.

Quantum theorems		
Trivial system	I	$\mathcal{H}_{\mathrm{I}} = \mathbb{C}$
Reversible transf.	$\mathcal{U} = U \cdot U^{\dagger}$	$U \in \mathbb{U}(\mathcal{H}_{\mathrm{A}})$
Determ. transformation	$\mathcal{T} \in \mathrm{Trn}_1(\mathrm{A} \to \mathrm{B})$	$\mathcal{T} \in \mathrm{CP}_{\leqslant 1}(\mathrm{T}(\mathcal{H}_{\mathrm{A}}) \to \mathrm{T}(\mathcal{H}_{\mathrm{B}}))$
Parallel composition	$\mathcal{T}_1 \in \mathrm{Trn}(\mathrm{A} \to \mathrm{B}),\ \mathcal{T}_2 \in \mathrm{Trn}(\mathrm{C} \to \mathrm{D})$	$\mathcal{T}_1 \otimes \mathcal{T}_2$
Sequential composition	$\mathcal{T}_1 \in \mathrm{Trn}(\mathrm{A} \to \mathrm{B}),\ \mathcal{T}_2 \in \mathrm{Trn}(\mathrm{B} \to \mathrm{C})$	$\mathcal{T}_2\mathcal{T}_1$
States	$\rho \in \mathrm{St}(\mathrm{A}) \equiv \mathrm{Trn}(\mathrm{I} \to \mathrm{A})$	$\rho \in \mathrm{T}^{+}_{\leqslant 1}(\mathcal{H}_{\mathrm{A}})$
	$\rho \in \mathrm{St}_1(\mathrm{A}) \equiv \mathrm{Trn}_1(\mathrm{I} \to \mathrm{A})$	$\rho \in \mathrm{T}^{+}_{=1}(\mathcal{H}_{\mathrm{A}})$
	$\rho \in \mathrm{St}(\mathrm{I}) \equiv \mathrm{Trn}(\mathrm{I} \to \mathrm{I})$	$\rho \in [0, 1]$
	$\rho \in \mathrm{St}_1(\mathrm{I}) \equiv \mathrm{Trn}(\mathrm{I} \to \mathrm{I})$	$\rho = 1$
Effects	$\epsilon \in \mathrm{Eff}(\mathrm{A}) \equiv \mathrm{Trn}(\mathrm{A} \to \mathrm{I})$	$\epsilon(\cdot) = \mathrm{Tr}_{\mathrm{A}}[\cdot E],\ 0 \leqslant E \leqslant I_{\mathrm{A}}$
	$\epsilon \in \mathrm{Eff}_1(\mathrm{A}) \equiv \mathrm{Trn}_1(\mathrm{A} \to \mathrm{I})$	$\epsilon = \mathrm{Tr}_{\mathrm{A}}$
Transformations as unitary interaction + von Neumann Luders		$\mathcal{T}_i\rho = \mathrm{Tr}_{\mathrm{E}}[U(\rho \otimes \sigma)U^{\dagger}(I_{\mathrm{B}} \otimes P_i)]$

Klasyczna teoria

Tabela.

Classical theory		
System	A	$\mathbb{R}^{n_{\mathrm{A}}}$
System composition	AB	$\mathbb{R}^{n_{\mathrm{AB}}} = \mathbb{R}^{n_{\mathrm{A}}} \otimes \mathbb{R}^{n_{\mathrm{B}}}$
Transformation	$\mathcal{T} \in \mathrm{Trn}(\mathrm{A} \to \mathrm{B})$	$\mathcal{T} \in \mathrm{Mrkv}_{\leqslant}(\mathbb{R}^{n_{\mathrm{A}}}, \mathbb{R}^{n_{\mathrm{B}}})$

Matematyczna aksjomatyzacja teorii klasycznej. Z każdym systemem A kojarzymy rzeczywistą przestrzeń euklidesową RnA. Ze złożeniem systemów A i B kojarzymy przestrzenie iloczynów tensorowych RnA RnB. Transformacje z systemu A do systemu B opisują substochastyczne macierze Markowa z przestrzeni wejściowej do przestrzeni wyjściowej. Reszta to proste wnioski wynikające ze szczególnych przypadków: przedstawiono je w tabeli .

Classical theorems

Trivial system	I	$\mathcal{H}_{\mathrm{I}} = \mathbb{R}$
Reversible transformations	$\mathcal{P}$	$\mathcal{P} \in \mathrm{Prm}(n_{\mathrm{A}})$
Transformation	$\mathcal{T} \in \mathrm{Trn}_{\leqslant}(\mathrm{A} \to \mathrm{B})$	$\mathcal{T} \in \mathrm{Mrkv}_{\leqslant}(\mathbb{R}^{n_{\mathrm{A}}}, \mathbb{R}^{n_{\mathrm{B}}})$
Teterm. transformation	$\mathcal{T} \in \mathrm{Trn}_{1}(\mathrm{A} \to \mathrm{B})$	$\mathcal{T} \in \mathrm{Mrkv}_{1}(\mathbb{R}^{n_{\mathrm{A}}}, \mathbb{R}^{n_{\mathrm{B}}})$
Parallel composition	$\mathcal{T}_1 \in \mathrm{Trn}(\mathrm{A} \to \mathrm{B})$, $\mathcal{T}_2 \in \mathrm{Trn}(\mathrm{C} \to \mathrm{D})$	$\mathcal{T}_1 \otimes \mathcal{T}_2$
Sequential composition	$\mathcal{T}_1 \in \mathrm{Trn}(\mathrm{A} \to \mathrm{B})$, $\mathcal{T}_2 \in \mathrm{Trn}(\mathrm{B} \to \mathrm{C})$	$\mathcal{T}_2\mathcal{T}_1$
states	$\mathbf{x} \in \mathrm{St}(\mathrm{A}) \equiv \mathrm{Trn}(\mathrm{I} \to \mathrm{A})$	$\mathbf{x} \in (\mathbb{R}^{n_{\mathrm{A}}})^{+}_{\leqslant 1}$
	$\mathbf{x} \in \mathrm{St}_1(\mathrm{A}) \equiv \mathrm{Trn}_1(\mathrm{I} \to \mathrm{A})$	$\mathbf{x} \in (\mathbb{R}^{n_{\mathrm{A}}})^{+}_{=1}$
	$p \in \mathrm{St}(\mathrm{I}) \equiv \mathrm{Trn}(\mathrm{I} \to \mathrm{I})$	$p \in [0, 1]$
	$p \in \mathrm{St}_1(\mathrm{I}) \equiv \mathrm{Trn}(\mathrm{I} \to \mathrm{I})$	$p = 1$
effects	$\epsilon \in \mathrm{Eff}(\mathrm{A}) \equiv \mathrm{Trn}(\mathrm{A} \to \mathrm{I})$	$\epsilon(\cdot) = \cdot\mathbf{x},\ \mathbf{0} \leqslant \mathbf{x} \leqslant \mathbf{1}$
	$\epsilon \in \mathrm{Eff}_1(\mathrm{A}) \equiv \mathrm{Trn}_1(\mathrm{A} \to \mathrm{I})$	$\epsilon = \cdot\mathbf{1}$