*Avant. Pismo Awangardy Filozoficzno-Naukowej*
wol. II, nr 2/2011 www.avant.edu.pl
ISSN: 2082-6710

# Ewolucja: inżynier systemów komputerowych projektujący umysły[52]

**Aaron Sloman**
przekład: Ewa Bodal i Nelly Strehlau

**Streszczenie**

To, czego w ciągu ostatnich sześciu lub siedmiu tego nauczyliśmy się na temat wirtualnej maszynerii w wyniku dużego postępu nauki i techniki, umożliwia nam zaoferowanie stanowisku darwinowskiemu nowej obrony przeciw krytykom, którzy twierdzili, że jedynie forma fizyczna – a nie zdolności umysłowe czy świadomość – może być produktem ewolucji poprzez dobór naturalny. Obrona ta porównuje zjawiska umysłowe, wspominane przez przeciwników Darwina, z treściami maszynerii wirtualnej w systemach obliczeniowych. Obiekty, stany, zdarzenia i procesy w owej maszynerii, które dopiero od niedawna umiemy projektować i konstruować, a które nie mogły być nawet przedmiotem rozmyślań w czasach Darwina, mogą współdziałać z maszynerią fizyczną, w której są zaimplementowane, nie będąc identycznymi ze swoją fizyczną implementacją, ani też nie będąc jedynie agregatami fizycznych struktur i procesów. Istnienie różnych rodzajów maszynerii wirtualnej (w tym zarówno „platformowych" maszyn wirtualnych, mogących hostować inne takie maszyny, na przykład systemy operacyjne, jak i „aplikacyjnych" maszyn wirtualnych, na przykład korektorów pisowni i gier komputerowych) zależy od skomplikowanych sieci połączeń przyczynowych związanych ze strukturami hardware'u i software'u, zdarzeniami i procesami, gdzie specyfikacja takich sieci przyczynowych wymaga konceptów, które nie mogą być zdefiniowane w kategoriach konceptów nauk

[52] Pierwsza wersja tekstu wystąpienia "Evolution of mind as a feat of computer systems engineering: Lessons from decades of development of self-monitoring virtual machinery" przygotowanego na Pierre Duhem Conference (Society for Philosophy of Science), ktora odbyła się 19 lipca 2011 w Nancy (Francja). Autor prosi o zaznaczenie, że niestety nie będzie mógł sam sprawdzić przekładu, choć bardzo by chciał…

fizycznych. Ta niedefiniowalność, a także możliwość zachodzenia różnych rodzajów automonitorowania w maszynerii wirtualnej, wydaje się wyjaśniać niektóre z rzekomo tajemniczych i nieredukowalnych cech świadomości, które motywowały krytyków Darwina, a także bardziej współczesnych filozofów, krytykujących sztuczną inteligencję. Wynikają z tego konsekwencje dla filozofii, psychologii, neurobiologii i robotyki.

Spis treści:



## 1. Wirtualne maszyny i przyczynowość

Czy zastanawialiście się kiedyś, w jaki sposób edytor tekstowy dopasowuje tekst, gdy dopisujecie nowy znak w już zapełnionej linijce? Jeśli dodatkowy znak powoduje, że długość linijki przekracza wyznaczoną szerokość tekstu, linijka zostaje przerwana i niektóre znaki z jej końca zostają przeniesione do początku następnej. To może spowodować podobny efekt w wielu kolejnych linijkach. Część wycięta z końca jednej lub więcej niż jednej linijki może zostać przeniesiona na kolejną stronę, będąc przyczyną tego, że ona z kolei również się przesunie. W niektórych przypadkach każda kolejna strona dokumentu zostaje zmieniona, choćby poprzez zmianę numerów stron. Podobne rzeczy mogą wydarzyć się, gdy korektor pisowni podczas pracy wykrywa błąd i poprawia go, wprowadzając lub usuwając znaki. Jeśli poprawione słowo staje się krótsze, może to spowodować skrócenie jednej lub więcej linijek, a czasem stron, w niektórych przypadkach przenosząc tekst z kolejnej linijki lub strony.

Mechanizmy wytwarzające takie zmiany są z reguły bardzo efektywne i niezawodne (w większości przypadków), jeśli chodzi o to, w jaki sposób zmiany zachodzące w dokumencie prowadzą do wymaganych efektów w innych jego częściach, jak również w róż-

nych częściach maszynerii fizycznej, takich jak aktywna pamięć, twarde dyski, wyświetlacze, a czasem nawet wydruki.

Wszystko to zależy od interakcji pomiędzy bardzo odmiennymi technologiami rozwijającymi się od połowy zeszłego stulecia, z których niektóre podlegają ciągłym zmianom (na przykład materiały oraz procesy budowy i projektowania wykorzystywane przy produkcji procesorów komputerowych, pamięci, interfejsów, sieci, wyświetlaczy oraz innych fizycznych komponentów. Niektóre komponenty mogą ulegać zmianie, gdy inne pozostają takie same, w wyniku pomysłowego wykorzystania specyfikacji interfejsów, które pozwalają programistom odpowiedzialnym za jeden komponent na ignorowanie zmian lub ich braku po drugiej stronie interfejsu. Dotyczy to na przykład sytuacji, gdy jedne rzeczy zmieniają się w szybkim tempie, podczas gdy inne pozostają bez zmian i dalej działają, na przykład projekt i wymagania programowe starego edytora tekstu (choćby takiego, jakiego sam używam), którego kod źródłowy pozostał identyczny przez niemal ćwierć stulecia.

Wynalezienie „chmury obliczeniowej" (w rzeczywistości: ponowne odkrycie przydatności mechanizmów używanych przez niektórych z nas od kilku dekad) może prowadzić do szerokiego geograficznego rozproszenia procesów, które w przeszłości zachodziły w jednym pudełku. Podczas gdy piszę ten tekst, zasiadając przed komputerem w moim pokoju w domu, używam edytora tekstów uruchomionego na innym komputerze w pewnej odległości, na moim wydziale. Edytor ten został wynaleziony we wczesnych latach osiemdziesiątych, chociaż dużo nowsze technologie (w tym technologie sieci komputerowych) pozwalają, by każdy przyciskany przeze mnie klawisz powodował – niemal natychmiastowo – zmiany w odległej maszynie, które następnie wywołują zmiany na wyświetlaczu mojego ekranu. Jeśli przełączę się na pracę na moim własnym komputerze, efekty są dla mnie z reguły nieodróżnialne, chociaż muszę wówczas zmienić swoje procedury zapisu kopii zapasowych i przenieść później stworzone pliki na maszyny uniwersytetu. Tak czy inaczej, nie ma różnicy w poprawnym opisie tego, co dzieje się z dokumentem, gdy dopiszę znak, a tego, gdy tekst zostanie zmieniony przez korektor pisowni.

Dzieje się tak, ponieważ dokument istnieje w uruchomionej wirtualnej maszynie, lub też – bardziej konkretnie – w egzemplarzu pewnego rodzaju wirtualnej maszyny, której egzemplarze mogą mieć bardzo różne fizyczne implementacje. Otworzenie zaawansowanych komputerów i obejrzenie ich z użyciem najbardziej wyszukanych z dostępnych czujników i urządzeń pomiarowych nie ujawni żadnych znaków, słów, stron czy błędów w pisowni, a już na pewno fabuły historii, bohaterów, teorii, debat, napomnień, dyskusji i tak dalej. Procesy ekstrakcji danych, służące wyszukiwaniu takich rzeczy, nie mierzą fizycznych właściwości używanych maszyn.

Co więcej, nawet ludzie projektujący oprogramowanie nie będą w stanie rozpoznać, że jest ono uruchomione na tych maszynach, poprzez otworzenie ich i obserwowanie wzorów fizycznej aktywności. Często podstawowe oprogramowanie zostało wynalezione na wiele lat przed technologią, która teraz zapewnia ich działanie, dzięki współpracy mię-

dzy projektantami języków programowania, kompilatorów, interpreterów, systemów operacyjnych, całego mnóstwa nowych urządzeń, sterowników urządzeń, protokołów sieciowych i postępowi w badaniach materiałowych i technologii elektronicznej.

Większość badaczy i inżynierów związanych z całą tą nauką i technologią postrzega tylko niewielki podzbiór w misternej sieci mechanizmów, do której się przyczyniają. Prawdopodobnie na chwilę obecną nikt na tej planecie nie obejmuje rozumieniem całości systemów, których używamy. Rzeczy działają, gdyż wiele różnych rodzajów maszynerii zostało zaprojektowanych tak, by móc razem pracować. Niektóre z nich zajmują się przesyłaniem energii, inne – pochłaniającymi energię fizycznymi zmianami w submikroskopowych komponentach, niektóre – fizycznymi połączeniami między podsystemami, jeszcze inne – przechowywaniem i transmisją informacji wyrażanej we wzorach bitowych (nie liczbach, jak się powszechnie uważa, gdyż wszystkie obliczenia numeryczne są przekazywane we wzorach bitowych, tak jak nienumeryczne procesy, na przykład tekstowa manipulacja).

Wiele elementów tej technologii zaprojektowanych zostało w taki sposób, by zachować związki pomiędzy częściami, podczas gdy zachodzą zmiany, lub by rozprzestrzeniać zmiany w strukturach informacji na bardzo szczególne sposoby. Propagacja struktur informacyjnych, na przykład wzorów bitowych lub bardziej skomplikowanych struktur implementowanych przez struktury bitowe, takich jak sentencje lub obrazy, różni się od rozprzestrzeniania energii, jaka ma miejsce na liniach transmisji mocy, czy od transportu materii, jaki zachodzi w rurach wodnych[53].

Przyczynowość we wzorach zmian struktur informacyjnych zarazem zależy i różni się od przyczynowości związanej ze zmianami w energii i materii (Bardziej szczegółową dyskusję na temat struktury informacji znaleźć można w (Sloman 2011)).

## 2. Warstwy maszynerii wirtualnej

W przypadku wprowadzenia zmiany w tekście komponowanego dokumentu, oprócz samego zdarzenia zmiany zachodzi też o wiele więcej zdarzeń w cyfrowych obwodach tworzących używany komputer. Miliony tranzystorów mogą zmienić stan, powodując przesunięcia wzorów bitowych w ich linearnym szeregu, wywołując inne zbiory zdarzeń modyfikujących wzorce magnetyczne na twardym dysku oraz sygnały wysyłane do ekranu, pokazujące, co się dzieje, lub w niektórych przypadkach kierujące systemem rozpoznawania mowy, zwykle nakłaniające namawiające? do dokonania dalszych zmian przy użyciu myszy i klawiatury.

Nie istnieje jednak żaden fizyczny linearny szereg wzorów bitowych. Ten szereg, podobnie jak dokument i mechanizmy oddziaływania na dokumenty, stanowi nie-fizyczną

[53] Znakomitą relację z części historii technologii informacyjnej, z włączeniem użycia sygnałów ogniowych wiele setek lat temu, znaleźć można w (Dyson 1997).

strukturę, którą przyjmują i na którą oddziałują mechanizmy stworzone przez projektantów systemu. Ta struktura może być traktowana jako linearna przez mechanizmy narzucające porządek jej częściom. Wszystkie te zmiany i przesunięcia wzorów bitowych są przeprowadzane i zależne od fizycznych struktur nawet na “niższym poziomie” i od procesów, w których tranzystory zmieniają swój stan, transmitowane są sygnały elektryczne, a energia jest zużywana i rozpraszana. Fizyk kwantowy mógłby opowiedzieć tutaj jeszcze inną, bardziej nawet skomplikowaną historię na temat zachodzących tu zjawisk. Co powiedzą fizycy przyszłych wieków, nie da się na razie stwierdzić.

Na wyższym poziomie abstrakcji linearny szereg obiektów tekstowych może zostać narzucony szeregowi wzorów bitowych bez bezpośredniego mapowania na przydzielony podzbiór wzorów. Na przykład implementacja może zawierać fragmenty tekstu rozproszone w „pamięci” tak, że struktura przechowująca każdą część zawiera informację na temat położenia następnej części (przy użyciu zmiennych wskaźnikowych pamięci lub adresów). Jeśli system przetwarzania tekstu jest w stanie użyć tych połączeń, może on traktować tekst jako linearnie uporządkowany. W tym przypadku maszyna wirtualna zajmuje się przetwarzaniem tekstu implementowanego w abstrakcyjnej maszynie wirtualnej, zajmującej się oddziaływaniem na wzory bitowe, które implementowane są w cyfrowych obwodach, co z kolei implementowane jest w molekułach, atomach i innych fizycznych komponentach.

Czy wynika z tego, że ktoś zajmujący się wynalezieniem korektora pisowni musi wiedzieć wszystko na temat przemieszczeń i zmian stanów miliardów cząstek subatomowych, lub milionów wzorów bitowych w systemach cyfrowych implementowanych w fizycznej maszynerii? Nie! Jeśli wymagana byłaby specyfikacja na poziomie szczegółów, zadanie to byłoby niemożliwe dla ludzkiego umysłu, choćby dlatego, że dokładnie taka sama zmiana w tekście w dokładnie tym samym miejscu w pierwotnym pliku, skutkująca dokładnie takimi samymi zmianami w układzie tekstu i bardzo podobnymi zmianami w tym, co widać na ekranie lub na papierze, mogłaby mieć miejsce przy użyciu zupełnie odmiennych transformacji wzorów bitowych i maszynerii fizycznej, nawet jeśli ten sam proces edycji powtórzony zostanie na tym samym komputerze w późniejszym czasie, na przykład w wyniku przypadkowego usunięcia już zedytowanej wersji pliku.

Zatem projektant korektora pisowni nie musi umieć myśleć o wszystkich możliwych zmianach w elektronicznych (lub, co gorsza, sub-atomowych) strukturach i procesach mogących zajść, gdy wykryty zostanie błąd pisowni, po to, by zapewnić, że właściwy wzór zmiany zostanie użyty za każdym razem by naprawić znaleziony błąd. Projektant korektora pisowni nie musi w ogóle znać się na fizyce.

Edytor tekstu z korektorem pisowni jest tylko jednym spośród wielu różnych typów maszyn przetwarzających informacje, który może zostać zaimplementowany jako maszyna wirtualna działająca na i z innymi maszynami, włączając w to być może kilka warstw maszyn na niższym poziomie. Jednym z ważnych faktów odnośnie wielu z tych wirtualnych maszyn jest to, że nie komputują one jedynie odpowiedzi na jakieś pytanie matematyczne czy logiczne, które to zadanie może zostać rozwiązane bez jakiejkolwiek inte-

rakcji ze środowiskiem. Przeciwnie, opisywane przeze mnie maszyny wirtualne mogą wchodzić w interakcje z elementami fizycznego otoczenia - jak na przykład, gdy korektor pisowni prosi użytkownika by wybrał spomiędzy możliwych opcji poprawy. Ta zdolność otrzymywania informacji z otoczenia i oddziaływania nań jest cechą wielu rodzajów systemów kontrolnych opartych na informacji, między innymi systemów kontroli lotów czy systemów przetwarzania informacji w robotach. W dalszej części przedstawię twierdzenie, że wszystkie żywe organizmy korzystają z informacji w swoich funkcjach kontrolnych, a bardziej wyrafinowane spośród nich wymagają w tym celu raczej korzystania z wirtualnej maszynerii implementowanej w fizycznych maszynach niż jedynie z samych maszyn fizycznych.

Jednym z osiągnięć Turinga była specyfikacja Uniwersalnej Maszyny Turinga (*Universal Turing Machine*, UTM), w której jakakolwiek inna maszyna Turinga mogła być emulowana poprzez sprecyzowanie jej cech na taśmie UTM (A. M. Turing 1936). Pomogło to przy dowodzeniu ważnych teorematów, np. odnośnie ekwiwalencji, rozstrzygalności i kompleksowości. Można ją też uznać za prekursora tego, co dziś nazywamy maszynerią wirtualną (której nie należy mylić z wirtualną rzeczywistością). Spróbuję pokazać jak kombinacja wirtualności, interakcji przyczynowej i (względnej) nierozstrzygalności może wyprodukować coś nowego w nauce. Następnych kilka sekcji objaśnia w bardziej ogólny sposób czym są wirtualne maszyny (a w szczególności maszyny nieopisywalne fizycznie, *non-physically describable machines* - NPDM[s]) i dlaczego mają one o wiele większe znaczenie w filozofii niż się to zazwyczaj uznaje oraz czemu niektóre z nich nie są ekwiwalentne żadnej z maszyn Turinga. Następnie przedstawię implikacje dotyczące ewolucji umysłu i świadomości.

## 3. Wirtualność

Idea UTM określiła, że maszyna obliczająca może zostać uruchomiona przez bycie zaimplementowaną jako *maszyna wirtualna* w innej maszynie. (Sądzę, że sedno tej myśli pojęła Ada Lovelace sto lat wcześniej.) Matematyczne właściwości trajektorii maszyny w jej stanie biernym nie będą zależały od tego, czy jest uruchomiona bezpośrednio w maszynerii fizycznej, czy jako maszyna wirtualna zaimplementowana w innej komputacji. Okazało się to niezwykle ważne dla teorematów meta-matematyki i informatyki oraz dla praktycznych okoliczności używania jednego komputera dla wielu celów, włącznie z dzieleniem czasu. Jedną z konsekwencji jest to, że maszyna Turinga implementująca inną maszynę Turinga może być również maszyną wirtualną zaimplementowaną w UTM: zatem warstwowe implementacje są możliwe. W następnych dekadach nowe wynalazki w inżynierii pojawiały się równolegle z odkryciami w matematyce z pewnymi konsekwencjami, którym nie poświęcono wiele uwagi, ale które są bardzo interesujące z punktu widzenia filozofii i potencjalnie posiadają również znaczenie w biologii. Zasugeruję później, że ewolucja biologiczna “odkryła” wiele zastosowań maszynerii wirtualnej na długo przed nami. Niestety, słowo “wirtualny” sugeruje coś “nierealnego” lub “nieistniejącego”, podczas gdy maszyny wirtualne mogą sprawić, że coś się

dzieje: mogą być przyczynami wielu skutków, w tym skutków fizycznych. W tym sensie są one, wraz z obiektami i procesami w nich zachodzącymi, rzeczywiste a nie wirtualne!

## 4. Przyczynowość i komputacja

Przyczynowość jest istotnym aspektem rozwoju inżynierii. Możliwe jest, na przykład, wziąć jakikolwiek skończony zbiór maszyn Turinga i emulować ich równoległe, synchroniczne działanie w UTM. Pokazuje to, że zsynchronizowany paralelizm nie produkuje żadnej jakościowo nowej formy obliczeniowej. Dowodami na to są teorematy dotyczące związków między abstrakcyjnymi strukturami matematycznymi, włącznie z sekwencjami stanów maszyn Turinga - i nie wspominające o przyczynowości fizycznej. Działająca maszyna fizyczna może być przypadkiem takiej abstrakcyjnej struktury matematycznej. Niemniej jednak, jako że jest ona fizyczna, wpływ mogą na nią wywierać czynniki fizyczne, np. powodujące zmianę jej prędkości. Sloman (1996) wskazał, że teorematy mogą zawodzić dla niezsynchronizowanych fizycznych maszyn Turinga. Na przykład, jeśli TM T1 powtarza output “0”, a T2 powtarza output “1”, i outputy te łączą się, by uformować sekwencję binarną, to jeśli coś powoduje, by szybkości T1 i T2 różnicowały się przypadkowo, i działają one cały czas, rezultat może być (i prawdopodobnie będzie) niemożliwą do wyliczenia nieskończoną sekwencją binarną, pomimo że zarówno T1 jak i T2 poddają się teorematom dotyczącym maszyn Turinga.

Podobnie jeśli maszyna posiada czujniki fizyczne i niektóre z jej operacji zależą od odczytów czujników, sekwencja wygenerowanych stanów może nie być możliwa do wyspecyfikowania przez żadną TM, jeśli jej środowisko nie jest równoważne TM. Zatem matematyczne teorematy limitujące nie stosują się do wszystkich fizycznie zaimplementowanych systemów przetwarzających informacje.

Matematyczne jednostki, takie jak liczby, funkcje, dowody i abstrakcyjne modele obliczeniowe nie posiadają lokalizacji czasoprzestrzennej, w przeciwieństwie do działających realizacji komputacyjnych, z włączeniem realizacji zlokalizowanych w sieciach. Podobnie nie ma związków przyczynowych a jedynie logiczne/matematyczne pomiędzy stanami TM, które stanowią dziedzinę teorii matematycznej, podczas gdy istnieją one w przypadkach działających, w zależności od użytej maszyny fizycznej i fizycznego otoczenia. Zatem pojęcia takie jak ‘niezawodność’ odnoszą się do tych fizycznych realizacji, ale nie do matematycznych abstrakcji. Z punktu widzenia matematyki nie ma różnicy między trzema osobnymi komputerami, na których uruchomiony jest dany program i jednym komputerem symulującym trzy komputery z uruchomionym programem. Niemniej jednak inżynier dążący do niezawodności wybierze trzy fizycznie odrębne komputery z systemem głosowania jako część systemu kontroli lotów zamiast matematycznie równoważnego, równie szybkiego skorzystania z jednego komputera (Sloman 1996).

Fizyczne szczegóły dzielenia czasu przez trzy maszyny mają konsekwencje przyczynowe. Kiedy trzy odrębne maszyny działające synchronicznie zmieniają jednocześnie

stany, nic nie dzieje się między stanami, podczas gdy w implementacji podczas dzielenia czasu na jednym komputerze, maszyna realizująca określone obliczenia musi przejść przez operacje, by zamienić jedną maszynę wirtualną na drugą. Takie procesy "zamian kontekstowych" mają pośrednie pod-stany, które nie mają miejsca w implementacji równoległej. Złośliwy intruz, lub niezłośliwy system operacyjny będzie miał możliwości interferowania z systemami dzielącymi czas w trakcie procesu zamiany kontekstowej, np. modyfikując emulowane procesy, przerywając je, lub kopiując i modyfikując ich wewnętrzne dane.

Takie możliwości interwencji (np. sprawdzanie, czy pod-proces nie narusza ograniczeń dojścia, lub transferu informacji między urządzeniami) są często używane zarówno w pojedynczych komputerach, jak i w komputerach połączonych w sieć i przyczynowo połączonych z otoczeniem zewnętrznym, np. fizycznych czujników i urządzeń kontrolujących, fabryk chemicznych, samolotów pasażerskich, klientów komercyjnych, systemów społecznych lub ekonomicznych, i wielu innych. W niektórych przypadkach urządzenia umożliwiające zmianę analogowo-cyfrową i cyfrowo-analogową i mechanizmy bezpośredniego dostępu do pamięci pozwalają teraz na ciągłą interakcję pomiędzy procesami. Zobacz też (Dyson 1997).

Na technologię wspierającą interakcje przyczynowe składają się (w kolejności przypadkowej): zarządzanie pamięcią, stronicowanie, zarządzanie cache, interfejsy różnych rodzajów, protokoły interfejsowania, konwertery protokołów, sterowniki urządzeń, procesory kodu procedury obsługi przerwania, algorytmy szeregowania, mechanizmy przywilejowania, mechanizmy kontroli zasobów, systemy zarządzania plikami, interpretery, kompilatory, systemy wykonawcze [run-time] dla różnych języków, programy odśmiecania pamięci, mechanizmy obsługujące abstrakcyjne typy danych, mechanizmy dziedziczenia, narzędzia do debugowania, kanały komunikacji wewnątrz i pomiędzy maszynami ("potoki" i "gniazda"), systemy dzielenia pamięci, firewalle, programy sprawdzające wirusy, wirusy software'u, systemy bezpieczeństwa, systemy operacyjne, systemy rozwijania aplikacji, serwery nazw, programy sprawdzające hasła, i więcej. Wszystko to wydaje się składać na skomplikowane sieci połączeń przyczynowych w działających systemach, włącznie z zapobieganiem, by coś miało miejsce, umożliwianiem, by pewne rzeczy miały miejsce w pewnych warunkach, zapewnianiem, że jeśli miejsce będą miały pewne rzeczy, inne rzeczy będą miały miejsce i w niektórych przypadkach czuwaniem nad mapowaniem pomiędzy procesami fizycznymi i wirtualnymi. Wspomniany wyżej przykład procesora tekstowego ilustruje jedną z prostszych sieci przyczynowych, które można znaleźć w systemach obliczeniowych.

## 5. Różnorodność wirtualnych maszynerii

Działająca maszyna wirtualna może mieć wiele skutków, włącznie z powodowaniem własnej zmiany. Zrozumienie, w jaki sposób maszyny wirtualne mogą sprawić, żeby cokolwiek się działo, wymaga potrójnego rozróżnienia pomiędzy (a) Modelami Matema-

tycznymi (Mathematical Models - MM), np. liczby, zestawy, gramatyki, dowody, etc., (b) Maszynami Fizycznymi (*Physical Machines* - PM), np. atomami, woltami, procesami chemicznymi, włącznikami elektrycznymi, itd., i (c) Działającymi Maszynami Wirtualnymi (RVM), na przykład kalkulatorami, grami, urządzeniami do formatowania czy udowadniania, korektorami pisowni, klientami pocztowymi, systemami operacyjnymi, itd.

MM to statyczne struktury abstrakcyjne, takie jak dowody i systemy aksjomatyczne. Podobnie do liczb, nie mogą nic zrobić. Rodzajem MM są działania maszyn Turinga, których właściwości są przedmiotem dowodów matematycznych. Niestety, niektóre użycia terminu "maszyna wirtualna" odnoszą się do MM, np. "wirtualna maszyna Javy" ("*the Java virtual machine*"). Są to abstrakcyjne, nieaktywne, matematyczne jednostki, nie zaś RVM, podczas gdy PM i RVM są aktywne i wywołują pewne działania.

Maszyny fizyczne na naszych biurkach potrafią teraz obsługiwać różne zbiory maszynerii wirtualnej różnymi rodzajami równocześnie połączonych komponentów, których siły sprawcze działają równolegle do sił sprawczych wirtualnych lub fizycznych maszyn, na których są one uruchomione i pomagają kontrolować te maszyny fizyczne. Część z nich to aplikacje RVM, które wykonują konkretne funkcje, np. grają w szachy, sprawdzają pisownię, wysyłają lub obsługują e-maile. Inne to platformy RVM, jak systemy operacyjne lub systemy wykonawcze języków programowania, które są zdolne obsługiwać wiele RVM wyższego poziomu. Różne RVM mają różne poziomy granularności i różne rodzaje funkcjonalności. Wszystkie różnią się od granularności i funkcjonalności maszynerii fizycznej.

Relatywnie proste przejścia w RVM mogą wykorzystać dużo większy zbiór zmian na poziomie kodu maszyny i nawet większy zbiór zmian fizycznych w używanej PM - dużo większy, niż człowiek jest w stanie pomyśleć. (Nie było to prawdą w odniesieniu do najwcześniejszych maszyn wirtualnych działających na komputerach jedno-procesowych z co najwyżej setkami lub tysiącami lokacji pamięci i brakiem interakcji zewnętrznych). Pomijając najprostsze programy, nawet specyfikacje kodów maszyny są nie do ogarnięcia dla ludzkich programistów. Mechanizmy automatyczne (w tym kompilatory i interpretery) są używane, aby mieć pewność, że procesy na poziomie maszyny wspierają pożądane RVM.

## 6. Automonitorowanie, autokontrola i automodyfikacja

W tym kontekście istnieją ważne różnice pomiędzy interpretowanymi i kompilowanymi językami programowania. Interpreter zapewnia *w sposób dynamiczny*, że przyczynowe związki wymienione w programie są podtrzymywane. Jeśli program zostanie zmieniony podczas działania, zachowanie interpretera (a w pewnych wypadkach zachowanie potencjalne) zmieni się. Dla porównania, kompilator *w sposób statyczny* tworzy instrukcje kodu maszynowego, by zapewnić systematyczne wypełnianie kolejnych zmian w programie, zaś pierwotny program nie ma w nim później żadnej funkcji. Zmienienie go nie

ma żadnych skutków, chyba że zostanie ponownie skompilowany (np. przy użyciu kompilatora inkrementalnego). Z zasady instrukcje kodu maszynowego mogą zostać zmienione bezpośrednio przez działający program (np. przez użycie komendy "poke" w Basic'u), ale jest to zazwyczaj możliwe jedynie dla relatywnie prostych zmian i nie byłoby prawdopodobnie odpowiednie do zmieniania skomplikowanego planu po odkryciu nowych przeszkód, zaś modyfikowanie fizycznych przewodów byłoby niemożliwe. Zatem automonitorowanie i automodyfikacja są najprostsze, jeśli wykonać je używając opisów procesów odpowiadających maszynie wirtualnej wysokiego poziomu wyspecyfikowanej w interpretowanym formalizmie, a najmniej wykonalne na poziomie struktury fizycznej. Monitorowanie procesów i modyfikowanie skompilowanych instrukcji kodu maszynowego stanowią przypadek pośredni.

Są dwie różne zalety używania odpowiedniej RVM: wcześniej wspomniana większa granularność zdarzeń i stanów w porównaniu z PM lub RVM niższego poziomu/warstwy oraz użycie ontologii związanej z domeną aplikacji (n. gra w szachy, kupowanie biletów lotniczych). Obie z nich są niezbędne w procesach projektowania, testowania, debugowania, rozwijania, a także automonitorowania w trybie wykonawczym i kontroli, które byłyby niemożliwe do specyfikacji na poziomie fizycznych atomów, molekuł czy nawet tranzystorów (z powodu wybuchowej kombinatoryki, zwłaszcza w wielo-procesujących systemach dzielących czas, gdzie mapowanie pomiędzy wirtualną i fizyczną maszynerią ciągle ulega zmianie). Mniejsza ziarnistość i ontologia oparta na aplikacji sprawia, że automonitoring jest bardziej praktyczny przy uruchomianiu programów interpretowanych na wysokim poziomie niż przy uruchomieniu programów kompilowanych przez kod maszynowy. Ma to związek z trzecim aspektem niektórych maszyn wirtualnych: ontologiczną nieredukowalnością.

## 7. Implementowalne ale nieredukowalne

Dwie prezentowane dotąd główne myśli są dość znane, to jest (a) VM może działać na innej maszynie (fizycznej lub wirtualnej), i (b) RVM (i ich komponenty) działające równolegle mogą współdziałać przyczynowo ze sobą nawzajem i z innymi elementami w środowisku. Trzecia konsekwencja dwudziestowiecznej techniki nie jest tak oczywista, mianowicie: niektóre VM zawierają stany, procesy i interakcje przyczynowe, których opisy wymagają pojęć niedefiniowanych w terminach języka nauk fizycznych: są to fizycznie nieopisywalne maszyny (NPDM). Wirtualna maszyneria może zwiększyć naszą ontologię rodzajów interakcji przyczynowych poza fizyczne interakcje.

Nie jest to forma mistycyzmu. Jest to związane z faktem, że teoria naukowa może wykorzystywać pojęcia (takie jak np. "gen", "walencja"), które nie są definiowalne w kategorii akcji i obserwacji, które mogą przeprowadzić naukowcy. Zaprzecza to zarówno "pojęciowemu empiryzmowi" filozofów takich jak Berkeley i Hume, pierwotnie rozgromionych przez Kanta (1781), a także jego współczesnej reinkarnacji, tezy "gruntowania

symboli" spopularyzowanej przez Harnada (1990), która również zakłada, że wszystkie pojęcia muszą pochodzić z określonych realizacji.

Alternatywna teza „dowiązania teorii" (*theory tethering*), wyjaśniona w (Sloman 2007), jest oparta na wniosku dwudziestowiecznej filozofii nauki, że niezdefiniowane symbole używane w głębokich teoriach naukowych czerpią swoje znaczenia przede wszystkim ze struktury danej teorii, chociaż formalizacja takiej teorii nie musi w pełni determinować, do czego dokładnie w świecie się odnosi, skoro każdy system formalny może mieć wiele różnych modeli (Tarskiego)[54].

Pozostała nieokreśloność znaczenia formalnie doprecyzowanej teorii jest po części zredukowana przez precyzujące formy obserwacji i eksperymentu (nazywane czasami „regułami łączenia" („*bridging rules*") lub „postulatami znaczenia" („*meaning postulates*") (Carnap, 1947)), które używane są w testowaniu i aplikowaniu teorii, „dowiązywaniu" semantyki teorii do konkretnych elementów lub aspektów świata. Znaczenia nigdy nie są jednoznacznie zdeterminowane, gdyż zawsze możliwe są nowe obserwacje i pomiary (np. ładunku elektronu), które zostaną zaadaptowane wraz z postępem naszej wiedzy i techniki.

Ontologie używane w specyfikacjach VM, np. pojęcia takie jak "pion", "zagrożenie", "przechwycenie" itp., używane w specyfikacjach VM do szachów, są również głównie definiowane przez ich role dla VM, których specyfikacje wyrażają eksplikacyjną teorię o szachach. Nie korzystając z tego rodzaju pojęć, niebędących częścią ontologii fizyki, projektanci nie mogą rozwinąć wszystkich implementacji, które obecnie okazują się użyteczne lub dostarczające rozrywki, zaś użytkownicy nie mogą zrozumieć do czego służy program ani z niego skorzystać.

Zatem gdy VM jest uruchomiona, jest również uruchomiona implementacja fizyczna (ze zmianami w stanie fizycznym, przemieszczeniami materii fizycznej oraz przejściem i rozproszeniem energii), lecz te dwie maszyny nie są identyczne: występuje pomiędzy nimi asymetryczna relacja. PM stanowi implementację VM, ale VM nie jest implementacją PM. Jest też wiele innych twierdzeń prawdziwych dla jednej z nich i fałszywych dla drugiej.

Uruchomiona VM do szachów może, w przeciwieństwie do PM, na której się opiera, zawierać zagrożenia i ruchy obronne. A zarówno „zagrożenie" jak i „obrona" nie mogą być zdefiniowane w języku fizyki[55]. Zatem nie wszystkie pojęcia użyte, by opisać obiekty, zdarzenia i procesy w RVM, są definiowane przez pojęcia fizyki, mimo że RVM implementowana jest na maszynie fizycznej. Szczegółowy opis PM nie jest specyfikacją VM, jako że VM byłaby taka sama gdyby implementowano ją na bardzo odmiennej ma-

[54] Na przykład wiadomo dobrze, że dowolny model aksjomatów geometrii rzutowej pozostaje modelem, jeśli linie i punkty są wymienione, a symbole predykatów i relacji zostają na nowo odpowiednio zinterpretowane.

[55] Ta kwestia wymaga omówienia szerszego, niż na to pozwalają ograniczenia rozmiarów tego tekstu.

szynie fizycznej z innymi procesami fizycznymi zachodzącymi podczas przeprowadzania choćby jednej konkretnej sekwencji ruchów w szachach.

Opis VM nie jest też równoważny względem jakiejkolwiek stałej rozdzielności opisów, jako że specyfikacja VM determinuje, jakie PM stanowią adekwatne implementacje. Programiści mogą popełniać błędy i błędy w wirtualnej maszynerii są wykrywane i usuwane, zwykle przez zmianę tekstowych specyfikacji abstrakcyjnej wirtualnej maszynerii, a nie maszynerii fizycznej. Kiedy naprawiony zostanie błąd w programie, nie musi on być różnie naprawiany w każdej fizycznej implementacji - różne kompilatory lub interpretery dla języka mogą poradzić sobie z mapowaniem pomiędzy wirtualną maszyną i fizycznymi procesami w różnych fizycznych maszynach i szczegóły te nie są częścią specyfikacji wspólnej maszyny wirtualnej.

Nie można również definiować stanów maszynowych i procesów VM na bazie fizycznych specyfikacji wejścia/wyjścia (jak to się zakłada w niektórych formach funkcjonalizmu), gdyż bardzo różne technologie mogą być użyte do implementowania interfejsów w tej samej wirtualnej maszynie, np. używając myszy, klawiatury, mikrofonu lub zdalnej poczty elektronicznej jako wejścia. Co więcej, niektóre VM wykonują o wiele pełniejsze zadania niż można to wyrazić w relacjach wejścia/wyjścia, np. system wizualny człowieka (lub przyszłego robota!) obserwującego gwałtowny nurt rzeki. (Porównaj z krytyką Skinnera w (Chomsky, 1959)).

Niedefiniowalność ontologii VM w kategoriach ontologii PM nie sugeruje, że RVM zawierają w sobie jakieś “duchowe coś”, które może istnieć niezależnie od fizycznej implementacji maszynerii, jak uznają ci, którzy wierzą w nieśmiertelne umysły lub dusze. Pomimo niedefiniowalności istnieją bliskie związki przyczynowe pomiędzy stanami VM i PM; to jednak zawiera takie zjawiska jak wykrycie zagrożenia powodujące wybór ruchu defensywnego, co jest procesem VM, który może wywołać zmiany w fizycznym wyświetlaczu i fizycznej zawartości pamięci. Możemy wobec tego mówić o tym, co czasem określa się “przyczynowością oddolną”, w dołączeniu do “przyczynowości odgórnej” i “przyczynowości bocznej” (*sideways causation*) w RVM, lub pomiędzy RVM działającymi równolegle.

Skomplikowany zestaw technologii hardware, firmware i software, rozwiniętej od czasów Turinga, umożliwił nam budowanie systemów przetwarzania informacji o ogromnej złożoności i wyszukaniu, wykonujących wiele zadań, które do tej pory były wykonywane jedynie przez ludzi, i pewnych zadań, których nawet ludzie nie potrafią wykonać. Jednak na dłuższą metę być może ważniejszy jest nowy sposób myślenia o fizycznie nieopisywalnej wirtualnej maszynerii z mocami przyczynowymi, który zaczęliśmy rozwijać. Nowe narzędzia konceptualne odnoszą się nie tylko do tego, co mogą zrobić ludzcy projektanci, ale także do tego, co auto-monitorujące i auto-kontrolujące systemy mogą być w stanie robić. Ma to głębokie znaczenie dla naszego rozumienia ewolucji.

## 8. Krytycy Darwina

Krytycy Darwina, część których cytowałem w (Sloman 2010a), argumentowali, że jego dowody potwierdzały jedynie hipotezę, że selekcja naturalna wytwarza fizyczne formy i zachowania. Żaden nie był w stanie zrozumieć, jak mechanizmy fizyczne mogą wytwarzać tajemnicze i zewnętrznie nieobserwowalne psychiczne stany i procesy (tzw. *luka eksplanacyjna*). Od czasów Darwina problem ten był kilka razy formułowany i nazywany na nowo, np. jako problem „świadomości fenomenalnej" (*phenomenal consciousness*, Block 1995) lub „trudny problem świadomości" (*hard problem of consciousness*, Chalmers 1996). Temat ten był poruszony, a później odsunięty na stronę przez Turinga (1950). Pozostaje niejasnym w jaki sposób genom może, w rezultacie procesów fizycznych i chemicznych, tworzyć problematyczne, najwyraźniej nie-fizyczne, zewnętrznie nieobserwowalne, osobiste doświadczenia (qualia) i procesy myślenia, czucia i pragnienia.

Wcześniej przedstawiłem Uniwersalne Maszyny Turinga jako teoretycznych prekursorów techniki, na której opierają się sieci interakcji uruchomionych maszyn wirtualnych (RVM), które wyczuwają i kontrolują rzeczy w swoim środowisku. Takie RVM są w pełni zaimplementowane w realizujących je maszynach fizycznych (PM), ale pojęcia używane, aby opisać stany i procesy w niektórych RVM (np. "pion" (pawn) i "zagrożenie" (threat) w szachowych VM) nie są definiowalne w języku nauk fizycznych. Obecnie rozwijamy biologiczne zastosowanie tych idei, wyjaśniając jak auto-monitorujące, automodyfikujące RVM mogą zawierać pewne cechy świadomości, takie jak qualia, uprzednio uważane za tajemnicze, wydeptując ścieżkę dla teorii dotyczącej sposobu ewolucji umysłu i świadomości.

## 9. Epigeneza: ciała, zachowania i umysły

Turing interesował się zarówno ewolucją, jak i epigenezą i poczynił pewne pionierskie sugestie dotyczące procesów morfogenezy - różnicowania się komórek w celu ukształtowania różnych części ciała podczas rozwoju. O ile wiem, nie pracował on nad tematem sposobu, w jaki genom może wytwarzać kompetencje całego organizmu, w tym zachowań o skomplikowanych strukturach warunkowych, gdzie podejmowane akcje zależą od wewnętrznej i zewnętrznej informacji zmysłowej, chociaż krótko rozważał uczenie się (Turing 1950)[56].

Zrozumiałym jest, że fizyczne zachowania, takie jak polowanie, jedzenie, uciekanie przed drapieżnikami i spółkowanie, powinny wpływać na biologiczne dostosowanie, i że ewolucja powinna wybierać mózg i inne modyfikacje, które wytwarzają korzystne zachowania. Istnieją jednak wewnętrzne nie-behawioralne kompetencje, których biologiczne zastosowania nie są tak oczywiste: myślenie, wspominanie, postrzeganie z przy-

[56] Jego sugestia nauki oparta na *tabula rasa* może być krytykowana.

jemnością, stwierdzanie, że coś jest dziwne i próby zrozumienia tegoż. Nie jest oczywistym, jak ewolucja biologiczna mogła wyprodukować mechanizmy zdolne do podtrzymania takich procesów psychicznych.

Wiele gatunków rozwija behawioralne i wewnętrzne kompetencje, które zależą od środowiska w trakcie rozwoju (np. to, którym językiem mówi dziecko i które problemy matematyczne są rozumiane), więc procesy kierowane przez genomy muszą tworzyć pewne od urodzenia sprecyzowane kompetencje, częściowo pod wpływem genomu, a częściowo pod wpływem kombinacji sygnałów sensoryczno-motorycznych podczas rozwoju (Held i Hein 1963; McCarthy 2008). Przynajmniej u ludzi wewnętrzne procesy kształtowania się kompetencji poprzez zmiany w mózgu muszą trwać na długo po narodzinach, co sugeruje, że genom kontynuuje produkowanie, powodowanie lub ograniczanie skutków (w tym zmian w seksualnych i rodzicielskich motywacjach i zachowaniach) na długo po tym, jak rozwinęły się główna morfologia ciała i mechanizmy sensoryczno-motoryczne.

## 10. Autotransformacja w biologicznych maszynach wirtualnych

Karmiloff-Smith (1992) przedstawia wiele przykładów, kiedy po osiągnięciu kompetencji behawioralnej w jakiejś dziedzinie, uczący się (włącznie z pewnymi gatunkami nie będącymi ludźmi) organizują na nowo swoje rozumienie danej dziedziny w taki sposób, który da im nowe możliwości myślenia i komunikowania się na temat tej dziedziny. Kiedy dzieci rozwiną kompetencje językowe oparte na znanych frazach, spontanicznie przestawiają się na używanie twórczej składni (*generative syntax*), która pozwala na znajdowanie rozwiązań nowych problemów, zamiast konieczności uczenia się empirycznie co działa, a co nie. Craik (1943) wykazał wartość takich mechanizmów w 1943, sugerując, że mogą być oparte na działających modelach mentalnych/psychicznych[57]. Grush (2004) i inni sugerują, że modele takie mogą działać jako symulacje lub emulacje. Jednak przy zastosowaniu do celów rozumowania, w przeciwieństwie do przewidywania statystycznego, wymagana jest rozkładalna struktura informacji, na przykład, gdy udowadnia się teorematy geometryczne (Sloman 1971).

Do modeli psychicznych, których używamy do wyjaśniania i przewidywania, zaliczają się takie obiekty jak koła zębate, rowery, obwody elektryczne i inne mechanizmy zbyt nowe, by być częścią historii naszej ewolucji. Zatem, przynajmniej w przypadku ludzi, proces budowania modeli nie może być w całości zakodowany w genomie: konkretne modele potrzebują informacji ze środowiska zdobytej po narodzinach oraz, w przypadku kreatywnych wynalazców, idei wymyślonych przez jednostki.

[57] Nie udało mi się ustalić, czy Craik i Turing mieli kiedykolwiek ze sobą styczność. Turing musiał mieć pojęcie o pracy Craika, był bowiem członkiem klubu Ratio, założonego na cześć Craika niedługo po jego śmierci w wypadku samochodowym w 1945.

Zatem genom specyfikuje nie tylko fizyczną morfologię i kompetencje fizycznego zachowania, ale także wielofunkcyjną architekturę przetwarzania informacji, rozwiniętą częściowo na sposób specyficzny dla danego gatunku, przez przedłużony okres czasu, częściowo pod kontrolą cech środowiska, i zawiera nie tylko mechanizmy interpretowania informacji sensorycznej i mechanizmy kontrolowania zewnętrznych ruchów, ale także mechanizmy budowania i uruchomiania predykcyjnych i eksplanacyjnych modeli struktur i procesów, bądź to znajdujących się w środowisku, bądź wymyślanych przez jednostkę[58]. W jaki sposób genom może specyfikować trwające procesy budowania/konstrukcji, by osiągnąć tę funkcjonalność? Nie sądzę, by ktokolwiek był bliski odpowiedzi, ale przypuszczam, iż ewolucja odkryła zalety maszynerii wirtualnej na długo przed ludzkimi inżynierami.

W poprzednich sekcjach wyliczyłem zalety wirtualnej maszynerii w systemach komputujących projektowanych przez ludzi i jej zalety w porównaniu z bezpośrednim specyfikowaniem, projektowaniem, monitorowaniem, kontrolowaniem i debugowaniem maszynerii fizycznej, wynikające z grubszej ziarnistości i wykorzystania semantyki odpowiedniej do aplikacji. Być może ewolucja biologiczna również uznała wykorzystywanie maszynerii wirtualnej za korzystne w specyfikowaniu rodzajów kompetencji na relatywnie abstrakcyjnym poziomie, unikając ogromnego skomplikowania przy specyfikowaniu wszystkich fizycznych i chemicznych szczegółów. Początkowa specyfikacja kompetencji behawioralnych w genomie może zajmować o wiele mniej miejsca i być prostsza do zbudowania lub wyewoluowania jeśli zastosować specyfikację maszyny wirtualnej, pod warunkiem, że inne mechanizmy zapewnią, że „język wysokiego poziomu" (*high level language*) jest we właściwy sposób mapowany w maszynerii fizycznej. Użycie procesów automonitorowania wymaganych przy nauce i modyfikowaniu kompetencji, w tym debugowanie ich, może być zupełnie niemożliwe do rozwiązania jeśli operacje/działania atomów, molekuł czy nawet pojedynczych neuronów są monitorowane i modyfikowane, lecz bardziej możliwe do rozwiązania, jeśli monitorowanie zachodzi na poziomie RVM.

Zatem coś w rodzaju kompilatora jest konieczne do podstawowych epigenetycznych procesów tworzących wspólne cechy w całym projektancie, a coś w rodzaju interpretera, aby prowadzić późniejsze procesy uczenia się i rozwoju.

## 11. Ewolucja organizmów z qualiami

Jak widzieliśmy, maszyneria wirtualna może zostać zaimplementowana w maszynerii fizycznej, a zdarzenia w maszynach wirtualnych mogą zostać przyczynowo połączone z innymi zdarzeniami w VM, a także z fizycznymi zdarzeniami zarówno w maszynie wspierającej, jak i w środowisku, w wyniku wykorzystania kompleksowej mieszanki

[58] W Sloman (1979, 2008) używam argumentu, że wymaga to rodzajów "języka" (w ogólnym znaczeniu słowa, włącznie ze zmiennością i semantyką kompozycyjną), które wyewoluowały, a u młodych ludzi rozwijają się, początkowo w celach wewnętrznego przetwarzania informacji, nie zaś zewnętrznej komunikacji. Możemy nazwać je „ogólnymi językami" (*generalised languages* – GLs).

technologii służącej generowaniu i podtrzymaniu wirtualno-fizycznych relacji przyczynowych, wytworzonych w ciągu ostatnich siedmiu dekad. Niektóre ze zdarzeń i procesów w maszynach wirtualnych nie są identyczne z daną maszynerią fizyczną, a opis tychże wymaga ontologii niedefiniowalnej w kategoriach ontologii fizyki.

Użycie takiej wirtualnej maszynerii może niezmiernie uprościć projektowanie, debugowanie, konserwację i rozbudowę kompleksowych systemów. Wreszcie – co może być najważniejsze, jeśli chodzi o maszyny, które muszą monitorować i modyfikować własne operacje – wykonywanie monitorowania i modyfikacji na poziomie wirtualnej maszynerii może być wykonalne, podczas gdy analogiczne zadania będą skomplikowane i mało elastyczne do wykonania, jeśli powierzyć je fizycznej maszynerii monitorującej i modyfikującej.

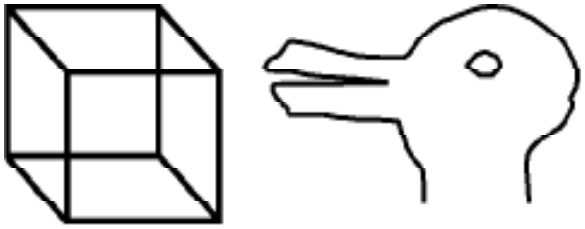
Zatem ewolucja biologiczna mogłaby zyskać na sile, elastyczności i prędkości rozwoju, gdyby korzystała z zawartych w genomie opracowań wirtualnych maszyn, by określać kompetencje behawioralne zamiast fizycznych szczegółów. Co więcej, jeśli część wirtualnej maszynerii nie jest w pełni doprecyzowana w genomie i musi być rozwinięta po narodzinach lub wykluciu przez użycie nowych informacji nabytych przez jednostkę z otoczenia, wówczas ów proces zachodzącej po narodzinach budowy będzie dużo prostszy w doprecyzowaniu, kontrolowaniu i modulowaniu, jeśli wykonywany będzie na poziomie maszyny wirtualnej, nie zaś przez doprecyzowanie wszystkich potrzebnych zmian chemicznych i neuronowych. Wreszcie: automonitorowanie, autokontrola i automodyfikacja w wyrafinowanym systemie przetwarzania informacji wymaga nie fizycznej, ale wirtualnej maszynerii.

## 12. W stronę rozumienia qualiów

Dla rozumnego organizmu postrzegającego, rozmyślającego i działającego w bogatym i skomplikowanym środowisku, zawierającym trwałe obiekty i procesy w różnych lokalizacjach, z których nie wszystkie są ciągle w zasięgu percepcji, korzystne będzie przechowywanie informacji o środowisku przy użyciu jednej lub więcej właściwych maszyn wirtualnych. Wzrokowe i dotykowe procesy percypowania tej samej części środowiska mogłyby składać się z nakładających się na siebie maszyn wirtualnych, zajmujących się różnymi aspektami środowiska, przetwarzanymi równolegle na różnych poziomach abstrakcji (Sloman 2009). Struktury danych, reprezentujące widoczne części i cechy środowiska, np. widoczne części powierzchni z kolorem, kształtem, ukierunkowaniem, krzywizną, szybkością ruchu lub obrotu oraz związkiem z innymi fragmentami powierzchni (a więc nie z konkretnymi sygnałami sensorycznymi), będą zatem komponentami maszyn wirtualnych. Jeśli do struktur informacji stworzonych podczas percepcji wzrokowej mają niekiedy dostęp automonitorujące procesy, które zajmują się nie tym, co znajduje się w środowisku, lecz treścią tego, co w chwili obecnej jest postrzegane, mamy potencjalne wyjaśnianie zjawisk, które doprowadziły do filozoficznych i nie tylko zagadek dotyczących istnienia i natury zmysłowych qualiów, które są czasem postrze-

gane jako coś, co definiuje najbardziej trudny do wyjaśnienia w funkcjonalnych kategoriach aspekt umysłu, i których ewolucję i rozwój w organizmach Huxley i inni uznali za tak trudne do wytłumaczenia. Zobacz też (Sloman i Chrisley 2003).

**Rysunek 1:**




*Rysunek 1: Każdy z tych dwóch obrazów jest wieloznaczny i zmienia się pomiędzy dwoma odmiennymi obrazami: (a) może być widziany jako trójwymiarowy druciany sześcian. Większość ludzi widzi na przemian dwa różne obrazy sześcianu, w których różnią się lokacje trójwymiarowe, orientacje i inne relacje. W przypadku (b) przemiana składa się ze zmian w częściach ciała, kierunku spojrzenia i prawdopodobnie ruchu – wymagając zupełnie innej ontologii.*

By zilustrować ten argument: kiedy wieloznaczne rysunki, np. Rysunek 1, są doświadczane jako zmieniające się z jednego widoku w drugi, wymaga to zmiany w treści pewnej maszynerii wirtualnej, a treść ta będzie reprezentowana na poziomie maszyny wirtualnej, odnoszącym się do różnych treści postrzegania, włącznie z odległością, kierunkiem nachylenia, częściami ciała, kierunkiem spojrzenia itd. Bardziej ogólnie: możemy próbować identyfikować treść informacji potrzebnej przy dużej różnorodności percepcyjnych doświadczeń, które pełnią pewną funkcję w umożliwianiu lub kontrolowaniu zachowania lub testowaniu teorii, generowaniu hipotez, niespodzianek lub pytań.

Przy próbach ponownego projektowania, testowania i debugowania, robocze przykłady takich procesów psychicznych zachodzących w robotach pomogą nam lepiej zrozumieć, jak można tak poszerzyć wirtualną maszynerię, by zawierała ona komponenty zdolne wykrywać, utrwalać i korzystać z informacji o tym, jaka jest zawartość doświadczeń percepcyjnych, czyli jakie są qualia. Zgodnie z argumentami podanymi w (Sloman i Chrisley 2003), jeśli pojęcia użyte dla zapisu takich metainformacji nie są w całości zaprogramowane z góry, lecz – jak zaproponowano w (Kohonen 1989) – generowane są przez wewnętrzne autoorganizujące klasyfikatory, wówczas wynikające zeń pojęcia, choć użyteczne dla jednostek bezpośrednio zaangażowanych, mogą być inherentnie nieprzekazywalne innym, gdyż ich użycie może *implicite* odwoływać się do mechanizmów dyskryminujących używanych w ich aplikacji, czego przykład Campbell (1994) nazywa “przyczynową indeksowalnością” (*causal indexicality*).

Ryle, Dennett i inni próbowali zidentyfikować niejasności pojawiające się w rozmowach o świadomości i qualiach, jednak rzeczy te z pewnością istnieją, mimo iż trudno je scha-

rakteryzować i zidentyfikować w innych jednostkach i gatunkach. Analiza przykładów, włączając w to dwuznaczne kształty, takie jak Rysunek nr 1, pomaga określić wymagania mechanizmów eksplanacyjnych. Obrazki takie ilustrują intencjonalność doświadczenia percepcyjnego, to znaczy interpretowanie czegoś jako odnoszącego się do czegoś innego oraz różne ontologie używane przy różnych doświadczeniach. Sugeruję, że jest to możliwe jedynie w ramach działającej maszynerii wirtualnej, jako że pojęcia takie jak "interpretowanie", "odwołanie", "zamierzanie" i "wyglądanie" są w języku fizyki lepiej definiowalne niż "pion" czy "zagrożenie".

Wiele organizmów potrafi, jak podejrzewam, tworzyć i wykorzystywać takie wirtualne byty, nie posiadając metasemantycznych mechanizmów koniecznych do wykrycia i wyrażenia tego faktu. Jednym z ważnych faktów odnoszących się do różnorodności rodzajów umysłu, wspomnianym w (Whittaker 1884) jest to, że nie wszystkie organizmy posiadające qualia wiedzą o tym! Możemy oddzielić występowanie treści psychicznej w organizmie od jej wykrycia przez organizm, co wymaga dodatkowej złożoności architekturalnej, mogącej podtrzymać mechanizmy autoobserwacji i autoopisu. Spodziewam się, że będziemy musieli eksperymentować z wyborem coraz bardziej skomplikowanych przykładów, używając różnych rodzajów mechanizmu, w celu lepszego zrozumienia pewnych pytań stawianych w kwestii zjawisk psychicznych w organizmach biologicznych. Jest to bliskie programowi badań Arbiba (2003).

## 13. Superweniencja, realizacja, tożsamość i warstwy?

Sekcja ta jest próbą połączenia pewnych sugestii poczynionych przez filozofów analitycznych na temat związku umysł-ciało i sposobu, w jaki można je odnieść do relacji między komputerami i ich obliczeniami. Temat ten jest bardzo szeroki i istnieje tu wiele szczegółowych rozważań; mogę jedynie wskazać na pewne różnice pomiędzy prezentowanymi tu ideami, a zestawem uprzednich prób scharakteryzowania związków umysł–ciało (Nie twierdzę, że zbadałem wszystko, co zostało napisane lub powiedziane na ten temat.)

Główną kwestią jest to, że, o ile wiem, żaden inny filozof nie próbował szczegółowo scharakteryzować związków pomiędzy zawartością działających maszyn wirtualnych w skomplikowanych systemach komputerowych i fizyczną technologią, od której zależą. O ile obliczenia są w ogóle wspominane w tym kontekście, zazwyczaj uznaje się je za albo (a) po prostu uruchomienie pojedynczego programu, który używa pewnych początkowych danych wejścia i potem wytwarza wynik (np. wyliczając wartość funkcji matematycznej dla pewnych argumentów lub odpowiadając na pytanie, czy dowód na konkretną formułę istnieje w pewnym systemie formalnym), albo (b) operację maszyny stanu skończonego, która w różnych momentach potrafi wybrać następne działanie na podstawie pewnego otrzymanych danych wejściowych, jak to opisuje na przykład Block (1996).

Systemy takie nie posiadają charakterystyk, które opisałem w działających maszynach wirtualnych, to jest wiele równolegle aktywnych nie-fizycznych (software'owych) mechanizmów, wpływających na swoje nawzajem zachowania, podczas gdy niektóre z nich są też połączone z konkretnymi wewnętrznymi podsystemami hardware'owymi, a także rzeczami dziejącymi się w środowisku, wyczuwanymi lub kontrolowanymi przez procesy w działających maszynach wirtualnych. System taki może być uważany za skomplikowaną sieć przyczynowo połączonych podsystemów, w których wiele funkcji kontrolnych i komunikacyjnych działa równolegle, z czego niektóre zawierają połączenia ze strukturami i procesami nie będącymi częścią systemu. Co więcej, podczas działania takiego systemu liczba podsystemów i połączenia miedzy nimi mogą ulec zmianie, tak jak może ulec zmianie podległa realizująca je maszyneria fizyczna, na przykład dlatego, że procesy są relokowane w pamięci maszyny, lub ponieważ pewne komponenty fizyczne są naprawiane lub wymieniane. Wiele systemów programowania zezwala nawet, by instrukcje programu zmieniały się w trakcie jego działania, poprzez zmianę interpretowanego kodu źródłowego lub używanie inkrementalnego kompilatora, który zmienia kod podczas działania[59]. To mogłoby umożliwić przyszłym robotom wytwarzanie swojej maszynerii wirtualnej podczas interakcji ze środowiskiem, tak jak zdaje się, że robią to niemowlęta i małe dzieci. Superweniencja maszyny wirtualnej jest o wiele bogatszym i głębszym związkiem niż superweniencja stanów lub właściwości. Ta pierwsza ma znaczące implikacje zarówno dla nauki, jak i filozofii umysłu i metafizyki.

Co więcej, kiedy w maszynach wirtualnych zachodzą zmiany, nie muszą one być wyrażane w mierzalnych jednostkach, takich jak rozmiar, orientacja, odległość, prąd, woltaż, pola magnetyczne, i tak dalej. Jest tak dlatego, że na procesy mogą składać się takie zjawiska jak konstrukcja, transmisja i analiza skomplikowanych strukturalnych bytów takich jak słowa, zdania, akapity, problemy, teorie, wyjaśnienia, diagramy, drzewa składni, wykresy, mapy topologiczne, formuły logiczne, dowody, intencje, plany, pytania, odpowiedzi na pytania, propozycje, decyzje i inne. (Niektóre z tychże to struktury noszące informacje, inne to treści informacji, część zaś należy do obu typów.) Dopóki wariacje w działających maszynach wirtualnych nie są w całości ilościowe, związki przyczynowe nie mogą być wyrażone formułami algebraicznymi, jak często zdarza się w naukach fizycznych i niektórych gałęziach inżynierii. Na przykład związki przyczynowe w korektorze pisowni lub wirtualnej maszynie grającej w szachy wyrażane są w formie algorytmów i baz danych, nie równań. Następstwem tego jest fakt, że duża część dyskusji filozoficznych na temat sposobu wykrywania związków przyczynowych przez porównywanie tempa zmian jest nieistotna dla tematu, skoro w wielu przypadkach nie ma dobrze zdefiniowanego pojęcia ilości zmiany, a zatem żadnej jednostki miary tempa zmiany - choć w niektórych przypadkach istnieją częściowe porządki - np. jeśli jeden zestaw zmian zawiera się w drugim, podczas gdy inne zestawy zaledwie się pokrywają. Zmiany takie muszą być opisywane raczej niż mierzone. Ma to konsekwencje dla kształtu teorii psychologicznych.

[59] Patrz: Poploghttp://www.cs.bham.ac.uk/research/projects/poplog/freepoplog.html

W przypadku pewnych treści maszynerii wirtualnej, takich jak treści skomplikowanych systemów wzrokowych u zwierząt, lub zmian, które zachodzą, kiedy matematyk patrzący na diagram zauważa sposób, w jaki można go zmodyfikować, by skonstruować dowód w geometrii Euklidesowej, nie wiemy jeszcze, czym są byty konstruowane i manipulowane przez różne psychiczne podsystemy. Stwierdzę tutaj bez podawania dalszych argumentów, że to, co dzieje się w większości przypadków przetwarzania wzrokowego u zwierząt pozostaje zagadką, chociaż znane jest wiele fizycznych i fizjologicznych szczegółów, włącznie z tym, które części mózgu są zaangażowane w te procesy.

Cała ta złożoność jest zazwyczaj ignorowana, kiedy filozofowie omawiają związki umysł-mózg. Na przykład jednym z częstych motywów w najnowszej filozofii jest dyskusja na temat tego, jak stany psychiczne lub właściwości psychiczne odnoszą się do/lub superweniują na stanach mózgu (lub, bardziej ogólnie, stany fizyczne, włączając w to aspekty środowiskowe). Jest to blady cień pytania, które zadaję na temat tego, jak skomplikowana maszyna psychiczna wykonująca zestaw funkcji postrzegania, nauki i kontroli, jest związana z i superweniuje na maszynerii fizycznej.

Jedna z ważnych idei wiąże się z pojęciem wprowadzonym przez G.E. Moore'a w odniesieniu do etyki około 1903. Powiedział on, że etyczne własności działań, takie jak ich dobro czy zło "superweniują" na ich własnościach pozaetycznych, takich jak: co zostało zrobione przez kogoś komu, z jaką intencją i jakimi konsekwencjami. Relacja superweniencji nie pozwala na logiczne wydedukowanie własności etycznych z pozostałych. Jest to słabszy związek, mianowicie niemożliwe byłoby odróżnienie dwóch działań etycznych, jeżeli nie różniłyby się co do własności pozaetycznych. Idea ta została przeniesiona na grunt filozofii umysłu przez D. Davidsona około roku 1970. Pytał on, czy psychiczne własności i stany superweniują na fizycznych własnościach i stanach w znaczeniu, że jest niemożliwym, by psychiczne własności lub stan danej osoby zmieniły się bez jednoczesnej zmiany fizycznej. Ta i związane z nią idee zostały rozwinięte przez różnych filozofów w ciągu ostatnich kilku dekad, na przykład przez Kima (1993, 1998).

Idea superweniencji jest przydatna, ale odnosi się do znacząco innych przypadków. Na przykład to, co opisywałem, można opisać jako „superweniencję maszyn wirtualnych" (*virtual machine supervenience*), jako że to, co superweniuje na systemie fizycznym lub niższy poziom wirtualnej maszynerii, nie jest stanem ani własnością, ale uruchomioną i zmienną maszyną wirtualną z podlegającymi interakcjom wewnętrznymi komponentami. Możemy to skontrastować z innymi rodzajami superweniencji.

„Superweniencja wzorca" (*pattern supervenience*) ma miejsce, gdy wzorzec zdefiniowany przez zbiór relacji istnieje z konieczności, gdy istnieje jakiś inny wzorzec: np. pionowe kolumny kropek superweniują na zbiórze poziomych rzędach kropek położonych w równych odstępach.

„Superweniencja zbiorcza" (*agglomerative supervenience*) (którą można by również nazwać "superweniencją część-całość") ma miejsce, gdy jakaś własność lub jednostka

definiowana jest poprzez kolektywny wkład wielu części obiektu. Do przykładów należą aspekty fizycznych obiektów, takie jak masa, środek grawitacji, moment pędu czy energia kinetyczna, które można wyliczyć z własności i ułożenia części. Są one czasem opisywane jako „użyteczne fikcje" przez filozofów, którzy źle zrozumieli ich rolę w nauce. Na przykład centrum grawitacji bryły sztywnej nie jest fikcją: jest to rzeczywiste miejsce zdefiniowane przez dystrybucję materii ciała. Siły skierowane przez centrum grawitacji (lub centrum masy) wywołują odmienne efekty niż inne siły. Zmiana centrum grawitacji może sprawić, że obiekt się przewróci.

„Superweniencja matematyczna" (*mathematical supervenience*) zachodzi, gdy zawsze, gdy coś ma własność P1, to ma też własność P2, ponieważ posiadanie P2 jest matematycznie wyprowadzalne z posiadania P1. Na przykład posiadanie nieparzystej liczby nóg może superweniować na posiadaniu pięciu nóg. Bycie wielokątem z pięcioma wierzchołkami superweniuje na byciu wielokątem z pięcioma bokami. W tym przypadku superweniencja jest symetryczna. Nie wiadomo, czy bycie sumą dwóch liczb pierwszych superweniuje na byciu liczbą parzystą.

Istnienie cieni ilustruje pewien rodzaj superweniencji. Cień nie może istnieć bez źródła światła, częściowo oświetlonej powierzchni i stojącego w drodze obiektu, który ten cień rzuca. Jeśli cień zmienia się w jakiś sposób, jakiś aspekt źródła światła, obiektu lub częściowo oświetlonej powierzchni również musiał ulec zmianie. Można to opisać jako „superweniencję przyczynową" (*causal supervenience*).

Przy pewnych rodzajach superweniencji istnieje dające się obronić twierdzenie, że superweniencja jest rodzajem tożsamości. Na przykład, jeśli istnieje wzór składający się z poziomych równo rozłożonych rzędów pełnych równo umieszczonych kropek, musi istnieć również wzór równo umieszczonych pionowych kolumn pełnych równo umieszczonych kropek. Można argumentować, że te dwa wzory są tym samym, postrzeganym różnie z powodu wybiórczej uwagi i opisanym w różny sposób.

Niektórzy filozofowie próbują rozwiązać zagadkę tego, jak stany mentalne mogą powodować zdarzenia fizyczne jeśli świat fizyczny jest przyczynowo zamknięty, argumentując, że mentalne stany, właściwości, zdarzenia, itd. nie tylko superweniują na bytach fizycznych, lecz są z nimi identyczne, a zatem przyczynowość mentalna po prostu jest tożsama z przyczynowością fizyczną.

To twierdzenie o tożsamości jest podważane przez istnienie procesów maszyny wirtualnej, których opisy wymagają pojęć (takich jak „atak" czy „zagrożenie" w szachach, „nieprawidłowa pisownia"), których nie można zdefiniować przy użyciu pojęć nauk fizycznych, wraz z twierdzeniem, że związek pomiędzy maszynerią wirtualną i maszynerią fizyczną nie jest symetryczny (nie superweniują wzajemnie na sobie). Nie ma tu jednak dość miejsca na pełną dyskusję.

Inny rodzaj próby odparcia twierdzeń, że zarówno fizyczne, jak i nie-fizyczne zdarzenia mogą być przyczynami, opiera się na twierdzeniu, że zdarzenia i procesy maszyn wirtu-

alnych są “epifenomenalne”, tj. są niezdolne do bycia przyczynami. Jednak w przypadku zdarzeń w wirtualnej maszynerii działającej w systemach obliczeniowych, jest to po prostu fałszywe: sednem projektowania i konstruowania wielu maszyn wirtualnych jest zapewnienie, że pewne dziejące się w nich rzeczy sprawiają, że mają miejsce inne rzeczy; istnieje obecnie ogromna ilość maszyn wirtualnych działających z powodu takich skutków, które bardzo trudno uzyskać przy użyciu jedynie maszynerii fizycznej. Ich produkcja w maszynerii wirtualnej jest również nietrywialna, lecz stało się to możliwe dzięki wielu staraniom i pomysłowości ze strony inżynierów zajmujących się hardware i software.

Nie twierdzę, że rozwiązałem tu wszystkie filozoficzne pytania i dysputy, choć mam nadzieję, że jasnym jest teraz, że opisane powyżej zjawiska superweniencji maszyn wirtualnych różnią się w znacznym stopniu od o wiele prostszych przypadków dyskutowanych wcześniej przez filozofów. Różnią się one swoją kompleksowością, bogactwem zawartości tego, co superweniuje, oraz mapowaniem pomiędzy fizycznymi maszynami i polegającymi na nich wirtualnymi maszynami; niekiedy zaś związki te zmieniają się raptownie. Próbowałem dowodzić, że było to bardzo ważne w ewolucji systemów przetwarzania informacji w organizmach, oraz wyraziłem przypuszczenie, że ewolucja wirtualnych maszyn mogących kontrolować, ewaluować, zapamiętywać i w inny sposób korzystać z części własnych zawartości maszyn wirtualnych (np. pośrednich form przedstawienia używanych w przetwarzaniu informacji wizualnych) koniec końców okaże się wyjaśnieniem dla zjawisk z których wynika filozoficzna dyskusja dotycząca qualiów. Możemy teraz doprecyzować, że qualia stanowią introspekcyjnie dostępne komponenty wirtualnej maszynerii i poprzez analizę funkcji wzroku możemy dowiedzieć się czemu istnienie auto-monitorujących mechanizmów z dostępem do zasobów qualiów jest, w pewnych sytuacjach, biologicznie przydatne

Pewien typ filozofów-funkcjonalistów próbuje zdefiniować różne rodzaje stanów umysłowych w kategoriach zestawów związków wejścia-wyjścia. Częstym i szeroko omawianym, zastrzeżeniem jest argument o zombie: anty-funkcjonaliści twierdzą, że mogą wyobrazić sobie zombie, czyli byty, których cechy zewnętrzne i widoczne zachowania we wszystkich okolicznościach sprawiają, że są nieodróżnialne od istot ludzkich, mimo że nie posiadają stanów i procesów umysłowych, a zwłaszcza świadomości. Nie wątpię, że wielu ludzi potrafi sobie to wyobrazić i w teorii maszyny takie mogłyby zostać zaimplementowane - zachowując się na ludzki sposób, lecz bez wewnętrznej maszynerii wirtualnej, którą opisuję. Mogłyby na przykład, przynajmniej w teorii, mieć wielkie tablice przeliczonych wcześniej danych (*lookup tables*), które determinowałyby zachowanie we wszelkich możliwych okolicznościach, zamiast tego rodzaju wirtualnej maszynerii rozpracowującej, co należy zrobić, o której pisałem.

Ale próby używania argumentu zombie przeciwko funkcjonalności wirtualnych maszyn, przedstawionego tu rodzaju, wymaga wyobrażenia sobie, że wszystkie wewnętrzne, niewidoczne, nie-fizyczne interakcje przyczynowe jednostki ludzkiej mogą zostać zreplikowane bez żadnych stanów czy procesów mentalnych w zombie pozbawionym świa-

domości, co jest zupełnie inną sprawą. Twierdzenie, że jest się w stanie to sobie wyobrazić to tylko filozoficzne przechwałki: żadna istota ludzka nie jest w stanie wyobrazić sobie wszystkich szczegółowych procesów maszyny wirtualnej koniecznych do zreplikowania ludzkiej funkcjonalności, a twierdzenie, że jest się w stanie wyobrazić, że wszystko to dzieje się w pozbawionym umysłu zombie nie powinno być brane poważniej, niż bezsensowne twierdzenie, że jest się w stanie wyobrazić, że cały wszechświat porusza się w kierunku zachodnim, lub twierdzenie, że wyobraża się, że w środku Ziemi panuje teraz południe. Ludzie mogą wyobrażać sobie, że to sobie wyobrażają, lecz nie dowodzi to, że naprawdę to sobie wyobrażają. W każdym razie historia matematyki pokazuje wyraźnie, że to, co ludzie myślą, że mogą sobie wyobrazić, nie jest dowodem prawdopodobieństwa.

## 14. Implikacje dla przyszłości filozofii

Doświadczenie poucza, że wielu myślicieli odrzuci tę dyskusję na temat wirtualnej maszynerii, której główne cechy nie mogą być opisane w języku fizyki, jako bezsensowne trele-morele, pomimo że właśnie taki rodzaj dyskusji okazuje się niezbędny dla projektowania i rozwijania coraz bardziej wyszukanych systemów przetwarzania informacji, używanych dla różnych celów praktycznych, i może być coraz ważniejszy, gdyż wymagamy od naszych systemów, by były coraz lepsze w posiadaniu wiedzy o tym, co robią i jak to robią. O ile wiem, jedyny proces edukacyjny mający na celu wytworzenie głębszego zrozumienia tych zagadnień polega na tym, że ludzie mają posiąść osobiste doświadczenia prób budowania maszyn, które potrafią robić to, co ludzie i inne zwierzęta, zamiast prób podjęcia dyskusji na te tematy tylko na podstawie ogólnych idei w kwestii komputacji.

Doświadczenie poucza także, że w przypadku myślicieli innego rodzaju nic z tej dyskusji nie wstrząśnie ich wiarą w istnienie nieprzekraczalnej luki eksplanacyjnej pomiędzy umysłem i ciałem. Jak argumentowałem w innym artykule (Sloman 2010b), niektóre przypadki protestów będą oparte na wykorzystaniu niespójnych pojęć (np. pojęcia „świadomości fenomenalnej”, zdefiniowanej tak, by nie zawierała żadnych mocy przyczynowych czy funkcjonalnych – zagadką pozostaje, jak wobec tego ludzie mają o nich rozmawiać). Mogą tu pomóc działające systemy, które ukazują, jak różne projekty robotów odnoszą się do różnych tworów ewolucji. Jest jednak prawdopodobne, że żaden rodzaj argumentu czy praktycznego doświadczenia w projektowaniu takich systemów nie przekona ludzi, którzy po prostu nie chcą uwierzyć, że działanie ludzkiego umysłu może zostać zrozumiane w kategoriach mechanizmów przetwarzających informacje, w tym być może mechanizmów działających w sposób bardzo różniący się od współczesnych komputerów.

Co więcej, aktualne osiągnięcia w dziedzinie Sztucznej Inteligencji (AI) dotyczące widzenia, kontroli motorycznej, formułowania pojęć, form uczenia się, rozumienia i używania języka, powstawania motywacji, podejmowania decyzji, formułowania planów,

rozwiązywania problemów oraz wiele innych są nadal (najczęściej) o wiele słabsze niż u ludzi i innych zwierząt, częściowo dlatego, że projektanci zazwyczaj rozważają tylko niewielki zestaw wymagań, typowych dla inteligencji biologicznej. Na przykład wyłącznym celem wielu badaczy jest stworzenie robotów z takim rodzajem kompetencji, którą Karmiloff-Smith (1992) określa jako "mistrzostwo behawioralne", ignorują zaś oni zupełnie inne rodzaje umiejętności rozwijanych przez człowieka, w tym umiejętność zastanawiania się z góry nad możliwymi kombinacjami działań, które są dostosowane do konkretnego środowiska i nigdy wcześniej nie były proponowane.

Jedną z umiejętności, które rozwijają młodzi osobnicy rodzaju ludzkiego, a także podgrupa innych gatunków, jest umiejętność uprzedniego zorientowania się, co się stanie w razie wykonania danego działania, zamiast dowiadywania się tego przez podjęcie próby takiego działania, która mogłaby okazać się śmiertelną, jak zauważa Craik (1943). Niektóre inne gatunki również wydają się posiadać ten rodzaj umiejętności. U ludzi zdaje się on być blisko związany z rozwojem kompetencji matematycznych, które pozwalają na rozwiązywanie problemów poprzez rozważanie abstrakcyjnych struktur, bez stałej potrzeby badania obiektów w środowisku, czego wymagają nauki fizyczne. Na chwilę obecną żaden z projektów przygotowywanych przez robotyków (o ile wiem) nie umożliwia robotom zauważenia matematycznych cech czasu, przestrzeni i ruchu, ani odkrycia czegoś takiego jak geometria euklidesowa, czy nawet podstaw topologii.

Nawet jeśli pominiemy kompetencje ściśle ludzkie, współczesne roboty są wciąż o wiele słabiej rozwinięte od innych zwierząt. Nie ma prostego sposobu, żeby zniwelować tę lukę, ale można podjąć się wielu prób, jeśli tylko zdamy sobie jasno sprawę z tego, co powinno zostać wytłumaczone.

Od pewnego czasu jest jasne dla filozofów, że nauka logiki jest konieczna do nauki filozofii matematyki, filozofii nauki, filozofii języka, a także epistemologii. Nadszedł czas, aby praktyczne doświadczenie w projektowaniu, testowaniu, debugowaniu, krytykowaniu i analizowaniu różnych funkcjonujących modeli ludzkiej i zwierzęcej kompetencji było postrzegane jako konieczny komponent w edukacji profesjonalnie kompetentnych filozofów umysłu, filozofów biologii i filozofów zainteresowanych aspektami metafizyki, zajmującymi się różnymi rodzajami przyczynowości[60]. Być może jednym ze skutków ubocznych takiej ekspansji edukacji filozoficznej będzie zbadanie potrzeby nowych form maszynerii przetwarzającej informacje fizyczne, konieczne do funkcjonowania tych aspektów przetwarzania informacji biologicznej, które nie zgadzają się z rodzajami maszynerii wirtualnej zdolnej do działania przy użyciu obecnej technologii komputerowej.

Podejrzewam na przykład, że potrzebujemy form przetwarzania informacji, które pozwalają na więcej rodzajów interakcji przyczynowych pomiędzy wirtualnymi maszynami, włącznie z nieustanną opozycją i konkurencją, czyli cechami, które obecnie przedstawiane są jedynie pośrednio, przy użyciu (do pewnego stopnia sztucznych) numerycz-

[60] Pochopnie uczyniłem to stwierdzenie dawno temu (Sloman 1978). Edukacja filozoficzna zmienia się jednak wolniej, niż tego oczekiwałem.

nych miar mocy lub ważności w wyborze między konkurującymi ze sobą alternatywami. Przez to nie można wyraźnie pokazać różnicy między wytworzeniem skutku konkurencji i przewidzeniem tego skutku. Obecnie różne komponenty maszyny wirtualnej nie mogą bezpośrednio przeciwstawiać się sobie nawzajem, choć mogą kontrolować fizyczne maszyny, które się sobie nawzajem przeciwstawiają, na przykład przez popychanie w różnych kierunkach. Dlatego trudne może być wytwarzanie dokładnych modeli sprzecznej motywacji lub niepowstrzymanych impulsów (na przykład impulsu do tego, by spojrzeć na coś w środowisku, lub by śmiać się, płakać, kichać czy drapać swędzące miejsce). Rozpocząłem badania nad takimi problemami, a na mojej stronie internetowej znajdują się niekompletne dyskusje na ten temat[61].

Jedno jest jasne: dopóki mówimy o maszynach, które zamiast tylko wyprowadzać struktury nowych informacji ze starych, pozostają też w ciągłej interakcji ze środowiskiem fizycznym i społecznym, a także kontrolują działanie różnych równoległych funkcji, takich jak chodzenie, mówienie, podziwianie widoków czy jedzenie kanapki – nie mówimy o maszynach Turinga, nawet jeśli niektóre z pomysłów Turinga są istotne dla tych zadań. Maszyny Turinga zostały zaprojektowane w celu badania możliwych form transformowania struktury informacji według ustalonych reguł. To, czego potrzebujemy obecnie, to maszyny projektowane (jak biologiczne systemy przetwarzania informacji) w celu posiadania różnorodnych funkcji kontrolnych.

Przypuszczam, że jak na razie ledwo zaczęliśmy badać niewielki podzbiór różnorodnych systemów przetwarzania informacji, które używane są przez rośliny i pewne zwierzęta. Niespodziewanie dużo czasu może zająć zaprojektowanie robotów, które uczą się i rozwijają tak jak ludzie, włącznie z rozwijaniem silnych zainteresowań oraz umiejętności matematycznych i filozoficznych. Bez wątpienia, kiedy je stworzymy, tak jak i my nie będą zgadzać się między sobą w temacie odpowiedzi na pytania filozoficzne, włączając w to pytanie, czy maszyny mogą mieć umysły.

**Bibliografia:**

Arbib, M. A. 2003. Rana computatrix to Human Language: Towards a Computational Neuroethology of Language Evolution. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 361 (1811): 2345-2379. Źródło: http://www.jstor.org/stable/3559127.

Block, N. 1995. On a confusion about the function of consciousness. *Behavioral and Brain Sciences*, 18, 227-47.

Block, N. 1996. What is functionalism? Źródło: http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/functionalism.html (Oryginalnie w: *The Encyclopedia of Philosophy Supplement*. Macmillan. 1996).

Campbell, J. 1994. *Past, Space and Self*. Cambridge: MIT Press.

[61] Http://www.cs.bham.ac.uk/research/projects/cogaff/talks/

Carnap, R. 1947. *Meaning and necessity: a study in semantics and modal logic*. Chicago: Chicago University Press.

Chalmers, D. J. 1996. *The conscious mind: In search of a fundamental theory*. New York, Oxford: Oxford University Press.

Chomksy, N. 1959. Review of skinner's Verbal Behaviour. *Language*, 35: 26-58.

Craik, K. 1943. *The nature of explanation*. London, New York: Cambridge University Press.

Dyson, G. B. 1997. *Darwin Among The Machines: The Evolution Of Global Intelligence*. Reading, MA: Addison-Wesley.

Grush, R. 2004. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27: 377-442.

Harnad, S. 1990. The Symbol Grounding Problem. *Physica D*, 42: 335-346.

Held, R., Hein, A. 1963. Movement-produced stimulation in the development of visually guided behaviour. *J. ofComparative and Physiological Psychology*, 56 (5): 872-876.

Kant, I. 1781/1929. *Critique of pure reason*. Przekład ang.: Norman Kemp Smith. London: Macmillan.

Karmilo-Smith, A. 1992. *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Kim, J. 1993. *Supervenience and Mind: Selected philosophical essays*. Cambridge: Cambridge University Press.

Kim, J. 1998. *Mind in a Physical World*. Cambridge, MA: MIT Press.

Kohonen, T. 1989. *Self-Organization and Associative Memory*. Berlin: Springer-Verlag.

McCarthy, J. 2008. The well-designed child. *Articial Intelligence*, 172 (18): 2003-2014.

Sloman, A. 1971. Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. W: Proc 2nd ijcai: 209-226. London: William Kaufmann. Źródło: http://www.cs.bham.ac.uk/research/coga/04.html#200407.

Sloman, A. 1978. *The computer revolution in philosophy*. Hassocks, Sussex: Harvester Press (i Humanities Press).

Sloman, A. 1979. The primacy of non-communicative language. Red. M. MacCafferty i K. Gray. The analysis of Meaning: Informatics 5 Proceedings ASLIB/BCS Conference, Oxford, March 1979:1-15. London: Aslib. Źródło: http://www.cs.bham.ac.uk/research/projects/coga/81-95.html#43.

Sloman, A. 1996. Beyond turing equivalence. Red. P. Millican, A. Clark. *Machines And Thought: The Legacy Of Alan Turing*, vol I: 179-219. Oxford: The Clarendon Press. (Presented at Turing 90 Colloquium, Sussex University, April 1990).

Sloman, A. 2007. Why symbol-grounding is both impossible and unnecessary, and why theory-tethering is more powerful anyway. Research Note No. COSY-PR-0705. Birmingham, UK. Źródło: http://www.cs.bham.ac.uk/research/projects/coga/talks/#models.

Sloman, A. 2008. Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking. *Generalised Languages*: GLs. Research Note No. COSY-PR-0702. Birmingham, UK.

Sloman, A. 2009. Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress. Red. B. Sendho, E. Koerner, O. Sporns, H. Ritter i K. Doya. *Creating Brain-like Intelligence*: 248-277. Berlin: Springer-Verlag.

Sloman, A. 2010a, August. How Virtual Machinery Can Bridge the Explanatory Gap. Red. S. Doncieux i in. *Natural and Articial Systems*. Proceedings SAB 2010, LNAI 6226: 13-24. Heidelberg: Springer.

Sloman, A. 2010b. Phenomenal and Access Consciousness and the “Hard” Problem: A View from the Designer Stance. *Int. J. Of Machine Consciousness*, 2 (1): 117-169.

Sloman, A. 2011. What's information, for an organism or intelligent machine? How can a machine or organism mean? Red. G. Dodig-Crnkovic, M. Burgin. *Information and Computation*: 393-438. New Jersey: World Scientic.

Sloman, A., Chrisley, R. 2003. Virtual machines and consciousness. *Journal of Consciousness Studies*, 10 (4-5): 113-172.

Turing, A. 1950. Computing machinery and intelligence. *Mind*, 59: 433/460 (przedruk w: E.A. Feigenbaum, J. Feldman. Red. *Computers and Thought.* McGraw-Hill, New York, 1963: 11-35).

Turing, A. M. 1936. On computable numbers, with an application to the Entscheidungsproblem. *Proc. London Math. Soc*., 42 (2): 230-265.

Whittaker, T. 1884, April. Review of G.J.Romanes Mental evolution in animals. *Mind*, 9 (34): 291-295.