



Ewolucjonista inteligencji

Wprowadzenie

Marcin Miłkowski

przekład: Krzysztof Ćwikliński

Trudno zaprawdę wskazać bardziej zagorzałego zwolennika tezy o ogromnej roli komputerów w filozofii niż Aaron Sloman. Jego książka z 1978 roku nosi tytuł *Computer Revolution in Philosophy (Rewolucja komputerowa w filozofii; Sloman 1978)*. Podkreślał w niej znaczenie informatyki dla rozumienia poznania:

*Może ona zmienić nasze myślenie o nas samych: dając nam nowe modele, metafory i inne narzędzia intelektualne, ułatwiające zgłębienie tajemnic ludzkiego umysłu i serca. Nowa dyscyplina **sztucznej inteligencji** jest gałęzią informatyki najbardziej bezpośrednio zaangażowaną w tę rewolucję. Dostarczając nam nowych, szerszych perspektyw na pewne nasze procesy wewnętrzne, zmienia nasze myślenie o nas samych, a więc również tę część naszych procesów wewnętrznych, przeistaczając więc i nas samych, tak jak każda rewolucja społeczna, techniczna i intelektualna.*

(Sloman 1978)

Jednak Sloman nie jest bynajmniej stereotypowym rzecznikiem sztucznej inteligencji. W swoich pracach podważa kilka popularnych przeświadczeń funkcjonalistów. Zwraca uwagę na fakt, iż uniwersalna maszyna Turinga (UTM) nie jest w istocie szczególnie ważna w modelowaniu poznania. Prawdziwe maszyny różnią się od abstrakcyjnych, a złożoność przyczynowa prawdziwych komputerów nie ma odzwierciedlenia w czysto abstrakcyjnych strukturach. Wskazuje, że model poznania bazujący na UTM ogranicza się do standardowych obliczeń cyfrowych – podczas gdy fizycznie możliwe jest, że dwie asynchroniczne maszyny Turinga, o ile tylko są połączone losowo, mogą obliczyć funk-

cje nieobliczalne przez UTM. Co więcej, Sloman nie używa standardowych argumentów funkcjonalistów, takich jak argumenty z wielorakiej realizowalności.

Mimo to Sloman daleki jest od standardowej krytyki sztucznej inteligencji: nie podąża tropem Searle'a, który głosi, iż komputery nie są w stanie osiąść prawdziwej intencjonalności; posuwa się nawet do twierdzenia, że poroniony w ogóle jest pomysł rozwiązywania problemu gruntowania symboli. Sloman, mówiąc oględnie, nie jest też zachwycony teoriami, które zaprzeczają roli reprezentacji w systemach poznawczych; krytykuje również radykalną wersję enaktywizmu, czyniącego z systemów poznawczych coś w rodzaju bardziej złożonych owadów.

Większość (jeżeli nie całość) prac Slomana jest dostępna na jego stronie internetowej, łącznie z licznymi prezentacjami, samouczkami, referatami i książką z 1978 r. Jak łatwo zauważyć, chętniej dyskutuje, niż pisze opasłe księgi; stwarza to pewien problem dla tych, którzy chcą cytować coś więcej niż tylko szkic ze strony. Tak czy inaczej, szkice i czasami bardzo ostre polemiki Slomana definitywnie wywierają dość istotny wpływ na filozofię sztucznej inteligencji.

Podczas konferencji CLMPS 2011 w Nancy miałem okazję usłyszeć wykład pt. *Evolution of mind as a feat of computer systems engineering: Lessons from decades of development of self-monitoring virtual machinery*, na którym opiera się artykuł Slomana. Na początku wykład wydawał się całkowicie trafny, ale nie mogłem zgodzić się z pewnymi тезami, jako że moja własna koncepcja implementacji obliczeń wychodzi z odmiennych założeń na temat przyczynowości i stosuje je jako podstawę wyjaśnień obliczeniowych (Miłkowski, w przygotowaniu). Artykuł Slomana przedstawia śmiałą hipotezę, że ewolucja ludzkiego umysłu polega w istocie na rozwoju wielu maszyn wirtualnych, które mają funkcje automonitorowania, co z kolei pomaga wyjaśnić szczególne własności naszego funkcjonowania poznawczego. Na marginesie zauważa kąśliwie, że współczesna filozofia analityczna nie rozpoznaje złożoności systemów przetwarzających informacje. Na przykład standardowe ujęcie superwencji wydaje się oparte na zbyt uproszczonych przypadkach, a także na naiwnym pojmowaniu przyczynowości. Stuprocentowa racja. Poniżej skupię się jedynie na tym, co uznaję za problematyczne, jako że tekst mówi sam za siebie i jest zbyt obszerny na szczegółowy komentarz. Są to uwagi raczej techniczne, ale uważam, że nie były wcześniej wystarczająco przedyskutowane.

Zgadzam się, że działające maszyny wirtualne naprawdę zwiększają złożoność komputerów, nie jestem jednak tak pewien, jak Sloman, że są one naprawdę tym, co odpowiada za automonitorowanie. Bez wątplenia pojęcie maszyny wirtualnej już od dawna postrzega się jako coś istotnego dla kognitywistyki, a Daniel Dennett podkreślał, że strumień świadomości może być czymś w rodzaju procesu w wirtualnej maszynie. Przeciwno tej idei wytoczono jednak istotne zarzuty:

W maszynach wirtualnych nie ma niczego szczególnego – w kontekście tej analogii (między maszynami wirtualnymi a świadomością – MM) – ani nicze-

go, dzięki czemu byłyby lepszym modelem umysłu niż jakakolwiek inna forma wykonywania oprogramowania. Wyobrażenie, że maszyna „może myśleć w obrębie własnej maszyny” nie jest bardziej przekonującym pomysłem niż jakiegokolwiek wykonywania programu, skoro maszyna wirtualna jest nie mniej osadzona w stanach maszyny od jakiegokolwiek innego procesu, który jest rzeczywiście zaimplementowany [...]. Wartość maszyny wirtualnej jako metafory opiera się w istocie na niezależności oprogramowania od sprzętu.

(Wilks 1992: 263)

Nie jest dla mnie oczywiste, czy można na ten zarzut łatwo odpowiedzieć: pojęcie maszyny wirtualnej w użyciu Słomana jest również tylko kolejną warstwą wykonywania programu.

Chociaż podkreśla on złożoność przyczynową wywołaną wieloma poziomami maszyn wirtualnych (VM), co jest oczywiście prawdą, pewne jego tezy na ten temat są problematyczne. Pierwszy problem polega na tym, że Słoman twierdzi, iż praca VM oddziałuje przyczynowo w sposób niefizyczny, co zdaje się dosyć radykalnym pomysłem. Dokładniejsza analiza wskazuje jednak, iż „niefizyczny” znaczy tyleż, co „niedefiniowalny w kategoriach fizycznych”; brak możliwości fizycznego zdefiniowania kategorii, w których zwykle opisuje się VM, jest tego samego rodzaju, jaki tradycyjnie przypisywano twierdzeniom o wielorakiej realizowalności. Innymi słowy, Słoman próbuje zapewnić teoretyczną autonomię kognitywistyki w stosunku do niższych, fizycznych poziomów przyczynowości. Zabrzmiałoby to dobrze w uszach nieredukcjonisty, ale znacznie trudniej to dziś obronić niż w czasach klasycznego funkcjonalizmu. Po pierwsze, klasyczna wizja redukcji, polegającej na logicznej derywacji teorii, opiera się na zdaniowym ujęciu teorii, które nie jest już dziś dogmatem w filozofii nauki. Po drugie, inna wizja redukcji, mianowicie mechanicystyczna, wydaje się znacznie bardziej rozpowszechniona w rzeczywistej nauce (patrz Bechtel i Richardson 1993).

Wyjaśnienie mechanistyczne, tj. wyjaśnienie funkcjonowania całych systemów w kategoriach przyczynowej organizacji ich części, opiera się na tego samego rodzaju związkach przyczynowych i nie wymaga definiowalności w kategoriach języka fizyki. Co istotne, dostarcza redukcjonistycznych wyjaśnień. Tak więc obrona autonomii na podstawie nieredukowalności w tradycyjnym, Naglowskim sensie, może się okazać nieszczerólnie dobrym pomysłem. Redukcja derywacyjna jest rzadkością i ze statystycznego punktu widzenia może być równie dobrze czymś, co można bezpiecznie pominąć, natomiast mechanistyczne wyjaśnienie jest codziennością w neuronaukach.

Jest jednak możliwe, iż moja interpretacja banalizuje twierdzenie o niefizycznej przyczynowości. Innym sposobem jego odczytania jest założenie istnienia szczególnego rodzaju przyczynowości, opartego na informacji. Słoman podkreśla, iż „nie wszystkie zmiany zachodzące w maszynach wirtualnych muszą być mierzalne ilościowo” i dodaje: „dzieje się tak, ponieważ te procesy mogą zawierać konstrukcję, transmisję i analizę złożonych elementów strukturalnych”. Najwyraźniej „niektóre wariacje w przebiegach

maszyn wirtualnych nie są mierzalne”, a zatem „relacje przyczynowe nie mogą być wyrażone wzorem algebraicznym”. Twierdzi więc, że takie stosunki przyczynowe nie są mierzalne, a jedynie opisywalne. Sęk w tym, że wystarczająco złożone koncepcje przyczynowości mogą sobie z tym łatwo poradzić i umożliwić testowanie zdań na temat przyczynowości metodami pomiarowymi (choć raczej nie redukcję wyłącznie do nich). Na przykład w koncepcji interwencjonistycznej używa się sieci Bayesa do modelowania stosunków przyczynowych (Pearl 2000; Spirtes, Glymour i Scheines 2001). Sieci Bayesa mogą być również z łatwością zastosowane do modelowania maszyn wirtualnych, jeśli tylko użyje się wystarczająco bogatego formalizmu, takiego jak ASM (Gurevich 1995). Jeśli nie odpowiadają nam formalizmy bazujące na przejściach stanów, można by posłużyć się abstrakcyjnym przepisywaniem napisów, chociaż można je równie łatwo modelować na strukturach grafopodobnych takich jak ASM. Zatem wciąż nie mam jasności, czy Slomanowi chodzi o wprowadzenie nowego pojęcia przyczynowości, czy też po prostu krytykuje zbyt jednostronne i uproszczone jej koncepcje.

Kolejną sprawą, o której Sloman wspomina na marginesie, jest przecenianie problemu gruntowania symboli; jego zdaniem, realny problem stojący u podłoża tego pierwszego został już właściwie dawno rozwiązany przez Carnapa. Jest to jedno z ulubionych twierdzeń Slomana, powtarzane latami; jednak do tej pory nikt z nim porządnie nie polemizował. Ale jego „rozwiązanie” jest jedynie pozorne. W czym rzecz? Problem gruntowania symboli jest pytaniem, jak symbole w systemach komputerowych mogą stać się reprezentacjami bez odwoływania się do żadnych zewnętrznych obserwatorów? Pozorne rozwiązania proponowane przez większość autorów, jak słusznie zauważa Sloman, są po prostu odmianami naiwnego empiryzmu pojęciowego. Te pomysły spalą więc na panewce, gdyż empiryzm pojęciowy jest niewiarygodny już od krytyk Kanta. To prawda. Propozycja Slomana jednak nie posuwa nas szczególnie naprzód. Zamiast gruntowania potrzebujemy „dowiązania”, powiada Sloman. Symbole powinny jedynie reprezentować na mocy podobieństwa strukturalnego, a choć podobieństwo nie wyznaczy przedmiotów desygnowanych przez te symbole jednoznacznie, to jednak tylko część znajdzie się w modelach standardowych teorii.

Są to modele dowiązane przez „zasady pomostowe”, które nie określają w pełni odwzorowania systemu symboli, a jedynie częściową redukują niedookreślenie ich znaczenia. Sęk w tym, iż „postulaty znaczeniowe” Carnapa czy „zasady pomostowe” nie są tak naprawdę obecnie przyjętym rozwiązaniem problemu teorii znaczenia w filozofii nauki, wbrew temu, co sugeruje Sloman. Chociaż opierają się one na pewnej wersji strukturalizmu, który jest wciąż w filozoficznym obiegu, korzystają z założeń niezgodnych z samą ideą dowiązania. Problematyczne jest, iż zasady pomostowe są zasadami, które odnoszą terminy teoretyczne do terminów obserwacyjnych. A więc terminy obserwacyjne są uznawane za znaczące same w sobie, a to właśnie założenie empiryzmu pojęciowego odrzucanego przez Slomana. Nie można mieć ciastka z Carnapem, nie jedząc również empiryzmu logicznego.

Bez terminów obserwacyjnych uznawanych za reprezentujące same w sobie dowiązanie nie zadziała; jednak te terminy nie są dowiązane do niczego same z siebie. Dla systemu zawierającego symbole terminy obserwacyjne Carnapa nie byłyby w ogóle znaczące. Byłyby tylko elementami kolejnego zbioru symboli. Jeśli owe terminy nie są znaczące *dla* systemu, są one zależne od obserwatora, a relacja reprezentowania opiera się na jego wiedzy, nie zaś na strukturze systemu, który właśnie używa symboli. Innymi słowy, skąd system wie, co *znaczą* terminy obserwacyjne użyte w zasadach pomostowych?

Propozycja Slomana nie jest żadnym rozwiązaniem problemu. Jest to po prostu kolejna wersja teorii reprezentacji jako odwzorowania: reprezentacja jest po prostu kwestią odwzorowania lub kodowania, które opierają się na współzależności bądź podobieństwie. Nawet gdyby było to wiarygodne dla reprezentacji zewnętrznej, nie może odegrać roli modelu reprezentacji umysłowej.

Taka właśnie koncepcja reprezentacji jako kodowania (*encodingism*), jak przekonująco ukazuje Mark Bickhard (Bickhard i Terveen 1995), nie jest wiarygodną teorią reprezentacji. Tylko jeśli symbol jest reprezentacją *dla* systemu, tj. kiedy odgrywa rolę w jego funkcjonowaniu jako reprezentacja, impas w problemie gruntowania zostanie przełamany. Są też inne warunki, które należy uwzględnić, czyli możliwość reprezentacji błędnej (związanej z realną intensjonalnością), jak również zdolność systemu do rozpoznania błędnej reprezentacji jako takiej. Przetwarzanie informacji czy stosunki przyczynowe nie uczynią z symbolu reprezentacji w pełnym sensie tego wyrazu, jeśli symbol nie odgrywa roli reprezentacji w systemie, nie tylko zaś w interpretacji obserwatora. Powinno być jasne dla Slomana, który tak wiele razy podkreślał, że realne komputery są systemami przyczynowymi o złożonej organizacji, których zachowanie jest nieredukowalne do czysto formalnego modelowania, i dla Slomana, który polemizował z koncepcjami obliczeń zrelatywizowanymi do obserwatora. Poprawna teoria reprezentacji musi opierać się na złożonej organizacji i dynamice przyczynowej w środowisku, a nie na obserwatorach zewnętrznych.

Bibliografia:

- Bechtel, W., Richardson, R. C. 1993. *Discovering complexity: Decomposition and localization as strategies in scientific research*. Discovery. Princeton: Princeton University Press.
- Bickhard, M. H., Terveen, L. 1995. *Foundational issues in artificial intelligence and cognitive science: Impasse and solution*. North-Holland.
- Gurevich, Y. 1995. Evolving algebras 1993: Lipari guide. E. Börger, red. *Specification and Validation Methods*: 231-243. Oxford: Oxford University Press.
- Miłkowski, M. (w przygotowaniu) *Explaining the Computational Mind*. Cambridge, MA: MIT Press / Bradford Book.

Pearl, J. 2000. *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press.

Sloman, A. 1978. *The Computer Revolution in Philosophy: Philosophy of Science and Models of Mind*. The Harvester Press. Available online at: <http://www.cs.bham.ac.uk/research/projects/cogaff/crp/>

Spirtes, P., Glymour, C. N. i Scheines, R. 2001. *Causation, prediction, and search*. Search. The MIT Press.

Wilks, Y. 1992. Dennett and Artificial Intelligence: On the same side, and if so, of what? A. Brooks i D. Ross. *Daniel Dennett*, Cambridge: Cambridge University Press.

Strona internetowa:

<http://www.cs.bham.ac.uk/~axs/>