

Wstęp

Ramy pojęciowe, które wnosimy do naszego badania poznania, mogą mieć ogromny wpływ na charakter tego badania. Stanowią zestaw filtrów, przez które patrzymy na świat, wpływając na wybór zjawisk do badania, język, w jakim te zjawiska opisujemy, pytania, jakie na ich temat zadajemy oraz interpretacje otrzymanych odpowiedzi. Przez większą część ostatnich pięćdziesięciu lat myślenie o myśleniu było zdominowane przez ramy obliczeniowe, czyli pogląd, że systemy są inteligentne w takim stopniu, w jakim potrafią kodować wiedzę w symbolicznych reprezentacjach, którymi następnie manipuluje się algorytmicznie w celu uzyskania rozwiązań problemów, które te systemy napotykają. Niedawno struktura koneksjonizmu wymusiła istotne udoskonalenie struktury obliczeniowej, w której reprezentację i obliczenia można rozłożyć na dużą liczbę jednostek luźno przypominających neurony. Mniej więcej od połowy lat 80., wraz ze wzrostem popularności koneksjonizmu, na scenie pojawiły się (lub, jak w przypadku koneksjonizmu, ponownie) kolejne ramy pojęciowe. Ramy te, które z braku bardziej chwytliwej etykiety będą nazywać ramami umiejscowionymi, ucieleśnionymi i dynamicznymi (SED), skupiają się na konkretnym działaniu i podkreślają sposób, w jaki zachowanie podmiotu wynika z dynamicznej interakcji pomiędzy jego mózgiem, ciałem i jego otoczenie. Spróbuję prześledzić historię poszczególnych wątków intelektualnych umiejscowionej aktywności, ucieleśnienia i dynamiki, które leżą u podstaw podejścia SED. Skoncentruję się szczególnie na latach 1985–1995. Chociaż istniały ważne prekursory podejścia SED (o niektórych pokrótce wspomnę), a prace w tej dziedzinie szybko rozwinęły się w ostatnich latach, wiele kluczowych idei zostało przedstawionych po raz pierwszy ich nowoczesną formę w tym dziesięcioletnim okresie.

Działalność zlokalizowana

Pierwszym wątkiem intelektualnym składającym się na podejście SED jest aktywność usytuowana. Z grubsza rzecz biorąc, aktywność usytuowana kładzie nacisk na trzy idee, które były tradycyjnie zaniedbywane w sztucznej inteligencji i kognitywistyce. S1 Konkretność działania. Właściwie podejmowanie działań w świecie jest bardziej fundamentalne niż abstrakcyjne opisy, które czasami o nim tworzymy. Chociaż świadome rozważanie niewątpliwie odgrywa swoją rolę, ostatecznym zadaniem inteligentnego agenta jest zrobienie czegoś, podjęcie konkretnego działania, którego konsekwencje wykraczają poza jego własną czaszkę. Położenie S2. Centralne miejsce zajmuje bezpośrednie otoczenie agenta i rola w jego zachowaniu. To środowisko jest nie tylko bogatym źródłem ograniczenia i możliwości dla agenta, ale także kontekst, w którym nadaje sens działaniom agenta. S3 Interakcjonizm. Relacja agenta ze swoim otoczeniem opiera się na ciągłej interakcji. Środowisko nie służy jedynie agentowi jako źródło izolowanych problemów do rozwiązania, ale raczej partner, z którym agent jest w pełni zaangażowany w improwizację z chwili na chwilę. Filozoficzne korzenie aktywności usytuowanej można doszukiwać się w fenomenologii, zwłaszcza w pracach Martina Heideggera (1927/1962), która została wprowadzona do sztucznej inteligencji i kognitywistyki przede wszystkim dzięki krytyce Huberta Dreyfusa (1972/1992). Jednym z kluczowych spostrzeżeń Heideggera było rozróżnienie, jakie dokonał pomiędzy przedmiotami będącymi *zuhanden* („gotowymi do ręki”) i *vorhanden* („obecnymi pod ręką”). W naszym codziennym doświadczeniu zwykle spotykamy rzeczy jako zasoby do natychmiastowego działania w służbie osiągnięcia naszych celów. Na przykład dla kogoś, kto wbija gwóźdź, młotek w pewnym sensie przestaje istnieć. Raczej, jak każde narzędzie, staje się jedynie przedłużeniem ramienia (czyli jest poręczne). Dopiero gdy wyraźnie przyjmujemy postawę intelektualną wobec młotka (np. z powodu złamania rękojeści i młotka nagle nie jest w stanie pełnić swojej normalnej funkcji), młotek wyłania się z nieartykułowanego tła rzeczy jako odrębny przedmiot charakteryzujący się ma swój własny zestaw właściwości (tj. staje się obecny pod ręką). Wielu autorów starannie artykułowało wyzwania, jakie idee fenomenologiczne stawiają przed kognitywistycznym światopoglądem, który zdominował myślenie w sztucznej inteligencji i

kognitywistyce, który nie tylko pojmuje poznanie jako zarządzaną regułami manipulację reprezentacjami symbolicznymi, ale także dokonuje zasadniczego rozróżnienia między fizycznym i psychicznym, pomiędzy ciałem a umysłem oraz pomiędzy środowiskiem a podmiotem działającym. Innym ważnym prekursorem aktywności usytuowanej była psychologia ekologiczna Jamesa Gibsona (Gibson 1979). Opierając się na swoich badaniach wzroku pilotów II wojny światowej, Gibson podkreślił strukturę właściwą środowisku organizmu oraz znaczenie relacji organizm/środowisko dla teorii percepcji. Na przykład sposób, w jaki zmienia się pole widzenia zwierzęcia podczas poruszania się w otoczeniu, niesie ze sobą wiele informacji na temat kierunku i prędkości ruchu, odległości do obiektów, orientacji powierzchni i tak dalej. Poglądy Gibsona ostatecznie obejmowały szeroko zakrojone odrzucenie kognitywizmu. Jednak dla naszych celów najważniejszym wkładem Gibsona jest jego koncepcja afordancji – możliwości działania, jakie otoczenie stwarza agentowi. Na przykład młotek Heideggera umożliwia wbijanie gwoździ ze względu na łatwość chwytania rękojeści oraz kształt i twardość główki. Co więcej, Gibson argumentował, że chociaż afordancje są dostrzegalnymi faktami dotyczącymi świata, są one ekologiczne w tym sensie, że ich znaczenie zależy od możliwości konkretnego organizmu. Na przykład otwór, który umożliwia przejście dla myszy, niekoniecznie umożliwia przejście dla człowieka. Trzeci ważny wpływ na działalność usytuowaną wywarły prace w naukach społecznych. Na przykład Lucy Suchman, antropolog badająca interakcję człowiek–maszyna, wyśledziła przerwy w komunikacji między człowiekiem a systemem pomocy dla kserokopiarki na podstawie błędnych założeń projektantów systemu na temat natury działania (Suchman 1987). Odrzuciła tradycyjny pogląd w sztucznej inteligencji i kognitywistyce, że działanie jest wynikiem wykonania planu, i zamiast tego argumentowała, że działanie należy rozumieć jako umiejscowione w tym sensie, że zależy od faktycznych okoliczności w miarę ich rozwoju. Z tego punktu widzenia wyraźne plany najlepiej interpretować jako zasoby służące komunikowaniu działań, a nie mechanizmy działania. Na podstawie swoich badań zespołu nawigacyjnego dużego statku morskiego inny antropolog, Edwin Hutchins podobnie doszedł do wniosku, że poznanie „na wolności” jest koniecznością często być rozumiane jako kulturowo ukonstytuowana aktywność grupy jednostek, zależna w dużym stopniu od rozwijającej się sytuacji, w której ma ona miejsce (Hutchins 1995). W obrębie sztucznej inteligencji idee usytuowane wysunęły się na pierwszy plan w połowie lat 80. Wcześniejsze demonstracje tego, jak bogate zachowanie może wynikać z prostych mechanizmów wchodzących w interakcję ze złożonym środowiskiem, obejmują roboty „żółwie” W. Graya Waltera (Walter 1953) i proste „pojazdy” Valentino Braitenburga (Braitenburg 1984). Jednakże badania aktywności umiejscowionej w obrębie AI powstały głównie jako reakcja na tradycyjne planowanie działania, w którym agenci reprezentują bieżącą sytuację i dostępne działania, formułują symboliczny plan działania, a następnie ten plan realizują. Philip Agre i David Chapman podkreślili niezdolność klasycznych technik planowania do zastosowania w złożonych, niepewnych środowiskach czasu rzeczywistego i zamiast tego zaproponowali, że rutynowe działania wynikają z interakcji prostych wewnętrznych mechanizmów z bezpośrednią sytuacją. Agre i Chapman wykazali użyteczność tego pomysłu w serii programów, z których najbardziej znanym był Pengi, agent, który grał w grę zręcznościową Pengo w czasie rzeczywistym, mimo że miał do czynienia z setkami często nieprzewidywalnych obiektów. Stanley Rosenschein i Leslie Kaelbling pokazali, jak specyfikację celów agenta można „zestawić” w prostą maszynę, tak że chociaż dla zewnętrznego obserwatora nadal miało sens mówienie o wiedzy i przekonaniach agenta, stany te nie odgrywały już żadnego bezpośredniego znaczenia. rolę w działaniach agenta. Wpływowa praca Rodneya Brooksa na temat robotów autonomicznych odrzuciła tradycyjny cykl zmysł-model-plan-działanie, podkreślając, że często „świat jest swoim najlepszym modelem”. Opracował warstwowy system sterowania znany jako architektura subsumpcji, w którym sieci prostych maszyn oddziałują ze sobą oraz z bezpośrednimi okolicznościami powodującymi zachowanie, i wdrożył go w różnych robotach. David Cliff i Beer wykazali znaczny potencjał interakcji pomiędzy pracą nad neuronalnymi podstawami zachowania zwierząt a czynnikami umiejscowionymi, opracowując modele odpowiednio

bzygowa i karalucha. Prawdopodobnie nikt nie zaprzeczy, że sytuacja środowiskowa odgrywa ważną rolę w zachowaniu agenta, ale fundamentalność tej obserwacji pozostaje kontrowersyjna. Niektórą działalność usytuowana ma posmak behawioryzmu, lecz zarzut ten w dużej mierze zależy od tego, co dokładnie rozumiemy przez „behawioryzm”. Z pewnością prawdą jest, że praca w działaniach umiejscowionych ponownie kładzie nacisk na konkretne zachowanie, a nie na abstrakcyjne rozumowanie. Jednakże abstrakcyjne rozumowanie nie jest odrzucane przez podejścia usytuowane, lecz raczej spychane do roli wspierającej jako ewolucyjnie niedawne opracowanie bardziej podstawowej zdolności poruszania się po świecie. Prawdą jest również, że wiele prac poświęconych aktywności usytuowanej miało tendencję do podkreślania architektur reaktywnych, w których działania podmiotu są całkowicie zdeterminowane przez jego doznania, oraz do odrzucania lub przynajmniej znaczącej rekonstrukcji idei wewnętrznych reprezentacji. Architektury reaktywne mocno przypominają paradygmat bodziec-reakcja przyjęty w behawioryzmie i mają dobrze znane ograniczenia, jeśli chodzi na przykład o zachowanie antycypacyjne. Jednakże, jak zobaczymy w dalszej części tej części, przywiązanie do architektur czysto reaktywnych jest niepotrzebne i możliwe jest wyartykułowanie roli stanu wewnętrznego, która jest zarówno istotna, jak i, co ciekawe, różni się od roli reprezentacyjnej, jaką taki stan odgrywa w tradycyjnej sztucznej inteligencji i kognitywistyka. Być może najbardziej kontrowersyjną koncepcją, która wyłoniła się w ostatnich latach z badań nad poznaniem umiejscowionym, jest koncepcja umysłu rozszerzonego (Clark 1997; Clark i Chalmers 1998). Pomysł ten opiera się na obserwacji, że nie tylko środowisko agenta odgrywa zasadniczą rolę w jego zachowaniu, ale sam agent może manipulować tą rolą poprzez aktywne organizowanie swojego otoczenia w celu zwiększenia swojej zdolności rozwiązywania problemów. Na przykład układamy składniki przepisu w kolejności, w jakiej będą potrzebne, i korzystamy z map, aby odnaleźć drogę przez rozległe miasta. Takie rusztowanie pozwala nam przenieść znaczną część naszego przetwarzania poznawczego na środowisko. Co więcej, za pomocą języka możemy koordynować działania wielu ludzi, tak aby mogli wspólnie dokonywać rzeczy, których nie byłaby w stanie dokonać żadna pojedyncza osoba, na przykład nawigować dużym statkiem morskim. Zwolennicy umysłu rozszerzonego argumentują, że jeśli pamięć, rozwiązywanie problemów itd. można rozłożyć na wiele czynników i artefaktów, to samo poznanie należy rozumieć jako zjawisko rozproszone, które wykracza poza czaszkę pojedynczego podmiotu i właściwie należy jedynie do większego systemu agentów i artefaktów, których ta osoba jest częścią. Rzeczywiście wiadomo, że nawet owady społeczne wspólnie wykonują złożone zadania konstrukcyjne, takie jak budowanie gniazd, modyfikując swoje środowisko w taki sposób, aby odpowiednio zorganizować przepływ pracowników i materiałów, co jest procesem określanym jako stygmergia

Ucieleśnienie

Drugim intelektualnym wątkiem w podejściu SED jest ucieleśnienie. Istnieją co najmniej trzy nieco odrębne idee, które zostały wysunięte przez zwolenników kognitywistyki ucieleśnionej.

E1 Fizyczne ucieleśnienie. Wyjątkowo fizyczne aspekty ciała agenta są kluczowe dla jego zachowania, w tym jego właściwości materialne, możliwości działania zapewniane przez układ i charakterystykę jego stopni swobody i efektorów, wyjątkową perspektywę zapewnianą przez szczególny układ i charakterystykę jego czujników oraz tryby interakcji sensomotorycznej, które wspólnie wspierają czujniki i efekторы. W pewnym sensie ten aspekt ucieleśnienia jest szczególnym przypadkiem usytuowania. Podczas gdy usytuowanie obejmuje każdy rodzaj interakcji ze środowiskiem, ucieleśnienie podkreśla te specyficznie fizyczne interakcje, w których pośredniczy ciało.

E2 Ucieleśnienie biologiczne. Ważne są nie tylko cechy fizyczne ciał, ale należy również wziąć pod uwagę specyficznie biologiczne fakty dotyczące istnienia organizmu, w tym odpowiednią neuronaukę, fizjologię, rozwój i ewolucję.

E3 Koncepcyjne wykonanie. Nawet gdy jesteśmy zaangażowani w czyste rozumowanie, nasze najbardziej abstrakcyjne koncepcje nadal są ostatecznie zakorzenione w naszych doświadczeniach cielesnych i metaforach zorientowanych na ciało.

Filozoficznych korzeni ucieleśnienia można doszukiwać się także w fenomenologii, zwłaszcza w pracach Maurice'a Merleau-Ponty'ego (1962), dla którego cielesne zaangażowanie w świat stało się centralnym elementem jego fenomenologii przeżywanego doświadczenia. Weźmy choćby jeden przykład: argument Merleau-Ponty'ego, że sposób, w jaki postrzegamy przedmiot, kształtowany jest przez rodzaje interakcji z nim, na jakie pozwala nasze ciało, można postrzegać jako wczesnego prekursora koncepcji afordancji Gibsona (1979). Myśl Merleau-Ponty'ego odegrała także ważną rolę w krytyce obliczeniowych teorii umysłu Dreyfusa (Dreyfus 1972/1992). W obrębie sztucznej inteligencji i kognitywistyki znaczenie fizycznego ucieleśnienia po raz pierwszy podkreślił Brooks (1991b). Brooks argumentował, że sztuczna inteligencja musi wyjść poza abstrakcyjne mikroświaty, które były jej głównym przedmiotem zainteresowania, i zacząć rozwiązywać problemy napotymane przez prawdziwe roboty poruszające się w rzeczywistych środowiskach. W ten sposób, jak zasugerował Brooks, stałoby się jasne, w jakim stopniu większość klasycznych technik sztucznej inteligencji jest po prostu nie do utrzymania w realistycznych sytuacjach. W łagodniejszej formie argument o fizycznym ucieleśnieniu polega po prostu na tym, że właściwości materialne ciała i środowiska odgrywają kluczową rolę w jego zachowaniu, a budując roboty, otrzymujemy tę fizykę „za darmo”, zamiast konieczności jej żmudnego modelowania. W najbardziej radykalnej formie twierdzi się, że tylko fizycznie utworzone systemy sztucznej inteligencji będą wykazywać naprawdę inteligentne zachowanie. W połączeniu ze współczesnymi trendami w poznaniu umiejscowionym, omówionymi w poprzedniej części, argumenty Brooksa wywołały eksplozję prac w dziedzinie robotyki opartej na zachowaniu, aktywnej percepcji, ucieleśniona kognitywistyka), autonomiczni agenci, niektóre aspekty sztucznego życia i filozofia umysłu. Ucieleśnienie biologiczne przenosi argumenty ucieleśnienia fizycznego o krok dalej. Ważne są nie tylko cechy fizyczne ciała, ale także biologiczne fakty dotyczące istnienia organizmu. Warunki niezbędne do utrzymania naszego stanu życia zasadniczo ograniczają nasze możliwości behawioralne i poznawcze. Ponadto specyficzne właściwości kości, mięśni i skóry, specyficzne cechy czujników biologicznych oraz sposób, w jaki te zdolności sensoryczne i motoryczne są powiązane w ludzkim ciele, zasadniczo definiują nasz własny, szczególny sposób ucieleśnienia. Co więcej, fakt, że przeszliśmy przez określoną historię ewolucji i rozwoju, może mieć również ważne konsekwencje dla naszej architektury behawioralnej i poznawczej. Na przykład Esther Thelen i Linda Smith argumentowały za znaczeniem zrozumienia sensomotorycznych początków poznania w rozwoju, zarówno w badaniach nad rozwojem chodzenia u niemowląt, jak i, ostatnio, w badaniach nad klasyczną pracą Jeana Piageta Błąd A-nie-B, w którym niemowlę wielokrotnie pokazuje przedmiot ukryty pod pudełkiem A, nadal będzie sięgać po A nawet po pokazaniu mu przedmiotu ukrytego pod drugim pudełkiem B. Podobny argument można wysunąć na rzecz wyłonienia się w ewolucji wyjątkowych ludzkich zdolności poznawczych od prostszych prekursorów. Wreszcie, nastąpił bardzo silny nacisk na włączenie większego realizmu neurobiologicznego do czynników ucieleśnionych. Z drugiej strony neuronauka zaczęła poważnie traktować rolę ciała i interakcji neuromechanicznych w kształtowaniu zachowań. Zatem konwencjonalne twierdzenie o ucieleśnieniu biologicznym jest takie, że cechy biologiczne organizmów mają znaczenie dla ich zachowania i poznania. Bardziej radykalne twierdzenie, czasami kojarzone z biologicznym ucieleśnieniem, głosi, że sam stan życia ma fundamentalne znaczenie dla poznania. Generalnie nie chodzi tu o to, że materiał czy biochemia życia jest niezbędna, ale raczej o to, że organizacja żywych systemów jest niezbędna dla ich zdolności poznawczych. Odpowiednie pojęcie żywej organizacji wywodzi się ogólnie z koncepcji autopoiezy Humberto Maturany i Francisco Vareli (w przybliżeniu samoprodukująca się sieć komponentów i procesów, tj. rodzaj homeostazy organizacyjnej). Wreszcie, pojęciowe ucieleśnienie dotyczy sposobu, w jaki nawet abstrakcyjne

koncepcje są często zakorzenione w cielesnym doświadczeniu i metaforze. Na przykład Stevan Harnad zdefiniował problem uziemienia symboli jako problem tego, w jaki sposób słowa, a ostatecznie stany mentalne, uzyskują swoje znaczenie i zaproponował, że sposobem rozwiązania tego problemu jest ugruntowanie ich w sygnałach sensomotorycznych. Co więcej, George Lakoff i Mark Johnson argumentowali, że struktura naszego rozumu opiera się na szczegółach naszego ucieleśnienia, a wiele abstrakcyjnych pojęć to metafory wywodzące się z dziedzin sensomotorycznych. Na przykład mówimy o zrozumieniu czegoś jako o „uchwyceniu” tego, a o niezrozumieniu czegoś jako o niepowodzeniu w „uchwyceniu” tego lub „przechodzeniu przez głowę”. Podobnie złe rzeczy „śmierdzą”, a „kawałki” teorii „pasują” do siebie.

Dynamika

Ostatnim wątkiem intelektualnym składającym się na podejście SED jest dynamika, w ramach której musimy wyróżnić co najmniej trzy idee.

D1 Teoria układów dynamicznych (DST). Teoria matematyczna, którą można zastosować do dowolnego układu charakteryzującego się stanem zmieniającym się w czasie w pewien systematyczny sposób.

D2 Ramy dynamiczne. Zbiór pojęć, intuicji i metafor związanych z przyjęciem dynamicznej perspektywy na jakiś system zainteresowań.

D3 Hipoteza dynamiczna. Specyficzna hipoteza wysunięta przez Timothy'ego van Geldera (1998) dotycząca sposobu DST i dynamiki ramy można połączyć w rygorystyczną kontrpropozycję wobec tradycyjnej hipotezy obliczeniowej w sztucznej inteligencji i kognitywistyce.

Układ dynamiczny to abstrakcja matematyczna, która jednoznacznie opisuje, jak stan jakiegoś układu ewoluuje w czasie. Składa się z przestrzeni stanów S , uporządkowanego zbioru czasu T i operatora ewolucji ϕ , który przekształca stan w pewnym momencie w inny stan w innym momencie. System dynamiczny, którego ewolucja zależy wyłącznie od jego stanu wewnętrznego, nazywa się autonomicznym, natomiast taki, którego ewolucja zależy również od czynników zewnętrznych, nazywa się nieautonomicznym. S może być numeryczne lub symboliczne, ciągłe lub dyskretne (lub hybryda tych dwóch) oraz o dowolnej topologii i wymiarze (w tym nieskończonym wymiarze). T jest zazwyczaj zbiorem liczb całkowitych lub zbiorem liczb rzeczywistych. Operator ewolucji może być podany jawnie lub zdefiniowany implicytnie i może być deterministyczny lub stochastyczny. Najczęstszymi przykładami układów dynamicznych są zbiory równań różniczkowych zwyczajnych i mapy iteracyjne, ale wiele innych rodzajów systemów matematycznych można również z powodzeniem opisać i przeanalizować w kategoriach dynamicznych. Dla każdego systemu matematycznego, który można ująć w tę postać, DST oferuje szeroką gamę narzędzi do analizy jego zachowania w czasie, z których wiele zostało opracowanych po raz pierwszy przez francuskiego matematyka Henriego Poincaré'a; na poparcie jego prac z zakresu mechaniki niebieskiej. Narzędzia te obejmują identyfikację zbiorów niezmienniczych (zbiorów punktów w przestrzeni stanów, których operator ewolucji nie zmienia, tj. punktów stałych i cykli granicznych), charakterystykę ich lokalnego zachowania (w jaki sposób reagują na zakłócenia, tj. ich stabilność) i ich globalne zachowanie (sposób, w jaki są ze sobą powiązane, tj. ich rozmaitości siodłowe) oraz ich zależność od parametrów (jak zmieniają się wraz ze zmianą parametrów, tj. ich rozwidlenia). Należy powtórzyć, że podobnie jak formalna teoria obliczeń, czas letni jest dziedziną matematyki, a nie naukową teorią świata przyrody. Pomimo tego, że czas letni nie jest sam w sobie teorią naukową, przyjęcie dynamicznej perspektywy na jakieś zjawisko naturalne niesie ze sobą zestaw pojęć, intuicji i metafor – pewien światopogląd – który wpływa na zadawane przez nas pytania, przeprowadzane przez nas analizy i jak interpretujemy wyniki. Kiedy podchodzimy do jakiegoś systemu z perspektywy obliczeniowej, skupiamy się na tym, jaką funkcję system próbuje obliczyć, w jakim formacie określone są dane wejściowe problemu, w jakim formacie

wyjściowym wymagana jest odpowiedź, w jaki sposób istotne cechy problemu mają zostać należy przedstawić, za pomocą jakich algorytmów te reprezentacje mają zostać przekształcone i w jaki sposób wydajność tych algorytmów skaluje się wraz z rozmiarem problemu. Z drugiej strony, gdy podchodzimy do jakiegoś systemu z perspektywy dynamicznej, staramy się zidentyfikować minimalny zestaw zmiennych stanu, których ewolucja może wyjaśnić obserwowane zachowanie, prawa dynamiczne, według których wartości tych zmiennych ewoluują w czasie, ogólną strukturę czasoprzestrzenną ich możliwej ewolucji oraz wrażliwości tej struktury na zmiany wejść, stanów i parametrów. Stwierdzono, że perspektywa dynamiczna jest owocna w wielu obszarach kognitywistyki. Dynamiczną perspektywę mózgu i zachowania po raz pierwszy wyraźnie sformułował W. Ross Ashby. W sieciach neuronowych Stephen Grossberg od dawna podkreśla znaczenie idei dynamicznych. Rzeczywiście, DST jest obecnie niezbędnym narzędziem w neuronauce obliczeniowej do analizy nie tylko pojedynczych komórek nerwowych lub małych obwodów, ale także całych systemów mózgowych. Idee dynamiczne zostały po raz pierwszy wprowadzone do psychologii ekologicznej przez Petera Kuglera (Kugler, Kelso i Turvey 1980; recenzje można znaleźć w Turvey 1990 i Warren 2006). Scott Kelso i współpracownicy od wielu lat zajmują się dynamicznym spojrzeniem na mózg i zachowanie, szczególnie podkreślając rolę samoorganizacji w tworzeniu wzorców zachowań i przejść między nimi. Thelen i Smith opowiadają się za dynamicznym podejściem do rozwoju poznawczego, w którym procesy i zmiany są badane przy użyciu tych samych narzędzi w różnych skalach czasowych. Jeffrey Elman podkreślił zasadniczo temporalny charakter rozumienia języka, w którym słowa poprzedzające silnie wpływają na interpretację kolejnych, i rozwinął dynamiczne podejście do języka. Na koniec argumentowałem, że teoria systemów dynamicznych zapewnia odpowiedni język teoretyczny i narzędzia do analizy rodzajów autonomicznych agentów, które były opracowywane w sztucznej inteligencji i robotyce (Beer 1995a), a Timothy Smithers (1995) i Gregor Schöner (Schöner, Dose i Engels 1995) opowiadali się za dynamicznym podejściem do projektowania robotów autonomicznych. Specyficznym sformułowaniem, któremu poświęcono wiele uwagi, jest hipoteza dynamiczna wysunięta przez van Geldera. Van Gelder definiuje system dynamiczny jako system ilościowy, to znaczy system, którego przestrzeń stanów, zbiór czasu i prawo ewolucji obejmują wielkości liczbowe. Jak widzieliśmy powyżej, jest to istotne ograniczenie matematycznej definicji układu dynamicznego. Jego hipoteza dynamiczna składa się zatem z dwóch elementów: (1) hipotezy natury i (2) hipotezy wiedzy. Twierdzenie hipotezy natury ma charakter ontologiczny: systemy poznawcze są systemami dynamicznymi. Natomiast hipoteza wiedzy głosi jedynie, że systemy poznawcze najlepiej zrozumieć za pomocą narzędzi teorii systemów dynamicznych. Biorąc pod uwagę, że nawet wielu zwolenników podejścia dynamicznego nie w pełni popiera dynamiczną hipotezę van Geldera, szkoda, że większość krytycznych dyskusji na temat dynamicznego podejścia do poznania skupiła się na specyficznym sformułowaniu van Geldera. Niemniej jednak jest to historycznie ważna próba sformułowania dynamicznej alternatywy dla hipotezy obliczeniowej.

W kierunku zintegrowanej perspektywy

Do tego momentu traktowałem usytuowanie, ucieleśnienie i dynamikę jako stosunkowo odrębne wątki intelektualne. Zrobiłem to zarówno dlatego, że historyczny rozwój tych idei nastąpił w pewnym stopniu niezależnie, jak i dlatego, że są one logicznie niezależne – to znaczy, że ludzie mogą i faktycznie traktują każdą z nich indywidualnie, niekoniecznie przyłączając się do pozostałych. Jednakże uważnemu czytelnikowi nie umknie fakt, że istnieje między nimi wiele potencjalnego nakładania się i synergii. Celem tej sekcji jest sformułowanie zintegrowanych ram teoretycznych, które łączą spostrzeżenia dotyczące umiejscowienia, ucieleśnienia i dynamiki. W przeciwieństwie do poprzednich części, w tej części przyjmę również bardziej osobisty punkt widzenia, opisując mój własny, szczególnie pogląd integracyjny, zamiast próbować ogólnego przeglądu wszystkich takich poglądów. Podstawowa

struktura usytuowana, ucieleśniona i dynamiczna (SED) jest dość prosta. Składa się z trzech następujących postulatów:

SED1. Mózgi, ciała i środowiska to systemy dynamiczne (por. S2, E1, E2, D1, D2). Wszystkie układy nerwowe, ciała i środowiska są konceptualizowane jako systemy dynamiczne, co mam na myśli jedynie to, że zakładamy, że każdy z nich można scharakteryzować poprzez zbiór stanów, których ewolucja w czasie jest rządzona przez prawa dynamiczne.

SED2. Dynamika mózgu, ciała i środowiska jest połączona (por. S1, S3, D1, D2). Układy nerwowe są zawarte w ciałach, które z kolei są umiejscowione w środowiskach, co prowadzi do gęstej interakcji pomiędzy tymi trzema układami składowymi. Sprzężony podsystem mózg-ciało będzie nazywany „agentem”. Sprzężenie, które przepływa ze środowiska do czynnika, będzie określane jako „zmysłowe”, a sprzężenie, które przepływa w przeciwnym kierunku, będzie określane jako „motoryczne”. „Zachowanie” agenta zostanie zdefiniowane jako trajektoria jego działań motorycznych.

SED3. Środek podlega ograniczeniom związanym z żywotnością (por. E2). Istnieją warunki dotyczące dynamiki agenta, które określają jego żywotność. Jeśli te ograniczenia żywotności zostaną naruszone, agent przestaje istnieć jako niezależna jednostka i nie może już angażować się w interakcje behawioralne ze swoim otoczeniem.

Teoretyczne zobowiązania aprioryczne dotyczące tych ram są dość minimalne. Rzeczywiście trudno sobie wyobrazić ramy teoretyczne, które nakładają mniej zobowiązań. Co może wyniknąć z tak małego zestawu twierdzeń? W rzeczywistości, jeśli poważnie potraktujemy te trzy postulaty, nastąpi niemal natychmiast wiele nietrywialnych konsekwencji. Być może najważniejszy wniosek jest następujący: ściśle mówiąc, zachowanie jest właściwością całego połączanego układu mózg-ciało-środowisko i ogólnie rzecz biorąc, nie można go właściwie przypisać żadnemu pojedynczemu podsystemowi w oderwaniu od pozostałych. Zdefiniowaliśmy zachowanie jako jedynie trajektorię działań motorycznych agenta. Jednakże, ponieważ dynamika mózgu, ciała i środowiska jest połączona, tworzą one jeden większy autonomiczny system dynamiczny z własnymi trajektoriami ewolucji czasowej. Trajektorie działań motorycznych podmiotu są jedynie projekcjami pełnych trajektorii całego układu mózg-ciało-środowisko i to właśnie te pełne trajektorie są właściwymi obiektami badań w ramach SED. Chociaż zachowanie jest właściwością całego sprzężonego układu, nadal warto zadać sobie pytanie o względny udział mózgu, ciała, i środowiska na jakąś szczególną cechę trajektorii behawioralnej. Aby to zrobić, musimy otworzyć sprzężony układ mózg-ciało-środowisko, przecinając jedną lub więcej ścieżek sprzężenia w celu wyizolowania komponentu, który chcemy zbadać. Składnik ten staje się wówczas nieautonomicznym systemem dynamicznym, a nasza analiza obejmuje zbadanie, w jaki sposób jego własna wewnętrzna dynamika oddziałuje z danymi wejściowymi, które otrzymuje od innych elementów sprzężonego systemu w procesie wytwarzania interesującej nas cechy behawioralnej. Ma to wiele interesujących konsekwencji dla sposobu, w jaki pojmujemy tradycyjne zjawiska behawioralne i poznawcze. Na przykład percepcję ogólnie postrzega się jako sposób, za pomocą którego podmiot wydobywa informacje o swoim otoczeniu z surowych sygnałów zmysłowych, które otrzymuje, i wewnętrznie reprezentuje strukturę swojego otoczenia. Jednak system dynamiczny podąża trajektorią określoną przez jego własny stan wewnętrzny i prawa dynamiki. Bodźce zmysłowe nie mogą na ogół wprowadzić nieautonomicznego układu dynamicznego w jakiś stan jednoznacznie charakterystyczny dla danego obiektu zewnętrznego. Zamiast tego jedyne, co mogą zrobić, to wpłynąć na wewnętrzne tendencje dynamiki agenta, wybierając jakąś konkretną trajektorię ze zbioru możliwych trajektorii, na które pozwalają prawa dynamiki agenta, z jego bieżącego stanu. Sugeruje to bardziej zorientowany na zachowanie pogląd na percepcję, przypominający Gibsona (1979). Z tego punktu widzenia percepcja to proces, w wyniku którego dynamika podmiotów odpowiednio wrażliwych na wpływy środowiska zostaje zakłócona przez trajektorię bodźców zmysłowych odbieranych przez system i przekształcanych

w zachowanie odpowiednie do okoliczności. Co więcej, ponieważ sprzężenie między agentem a jego otoczeniem jest dwukierunkowe, działanie agenta może kształtować jego własną percepcję. Agenci nie tylko postrzegają, aby działać, ale także działają, aby postrzegać. Ponieważ agenci w ramach SED są dynamiczni, nie są podatni na krytykę kierowaną pod adresem agentów reaktywnych. Czynniki reaktywne to takie, których moc motoryczna zależy wyłącznie od bodźców zmysłowych; jest to jedynie funkcja przechodząca od wrażenia do działania. Choć taki agent może uczestniczyć w złożonych interakcjach, gdy jest połączony z dynamicznym środowiskiem, jego zachowanie jest zawsze podporządkowane temu środowisku, ponieważ nie posiada własnej dynamiki. Natomiast odpowiedź czynnika dynamicznego jest zdeterminowana przynajmniej częściowo przez jego własną dynamikę wewnętrzną. Ponieważ posiada stan wewnętrzny, czynnik dynamiczny może różnie reagować na ten sam bodziec zmysłowy w różnym czasie, może inicjować zachowanie niezależnie od jego bezpośredniego środowiska, może modyfikować swoje zachowanie w oparciu o historię interakcji i może wykorzystywać długoterminowe korelacje w swoim otoczeniu, aby zorganizować swoje zachowanie w oczekiwaniu na przyszłe wydarzenia. Istotną zaletą modelu SED jest to, że oferuje on możliwość jednolitego traktowania odmiennych zjawisk behawioralnych i poznawczych, które często były postrzegane jako nie do pogodzenia. Z jednej strony niektóre podstawowe zachowania sensomotoryczne mogą mieć głównie charakter reaktywny, a stan wewnętrzny odgrywa jedynie niewielką rolę w „kolorowaniu” reakcji podmiotu na otoczenie. Z drugiej strony, niektóre z naszych najbardziej poznawczych zachowań można uznać za niemal oddzielone od bezpośrednich okoliczności środowiskowych, napędzane głównie przez czasową ewolucję stanu wewnętrznego. Oczywiście większość zachowań jest zazwyczaj mieszaniną wpływów zewnętrznych i wewnętrznych, przy czym względne znaczenie tych dwóch czynników zmienia się, czasem znacznie, z chwili na chwilę. Rzeczywiście, interesujące pytania dotyczące tego, w jaki sposób wyższe procesy poznawcze powstały z bardziej podstawowych kompetencji sensomotorycznych w trakcie ewolucji i rozwoju, wydają się znacznie bardziej przystępne w ramach teoretycznych, które stawiają je na wspólnej podstawie. Z tego punktu widzenia wyższe poznanie niekoniecznie zmienia nasz zasadniczo umiejscowiony, ucieleśniony i dynamiczny charakter, ale zamiast tego wzmacnia go znacznie zwiększonym zasobem wewnętrznej dynamiki. Jak mamy rozumieć naturę i rolę tego stanu wewnętrznego w czynniku dynamicznym? Tradycyjna interpretacja obliczeniowa takich stanów byłaby reprezentacjami wewnętrznymi. Jednak posiadanie stanu wewnętrznego jest właściwością układów fizycznych w ogóle i stany te mogą współzmienniać się ze stanami poza układem w dość skomplikowany sposób. Chyba że chcemy nadać status reprezentacji wszystkim stanom fizycznym (czy burza reprezentuje topografię terenu nad którą przechodzi?), muszą istnieć dodatkowe warunki, które uprawniają modyfikator do „reprezentacyjnego”. Niestety, pomimo fundamentalnej roli, jaką pojęcie reprezentacji odgrywa w podejściach obliczeniowych, istnieje bardzo niewielka zgoda co do tego, jakie mogą być te dodatkowe warunki. Rozważania te skłoniły mnie do przyjęcia stanowiska reprezentacyjnego sceptycyzmu (a nie, jak niektórzy sugerują, antyreprezentacjonizmu). Reprezentacyjny status państwa wewnętrznego postrzegam jako kwestię empiryczną, którą należy rozstrzygnąć zgodnie z precyzyjną definicją konkretnego oferowanego pojęcia reprezentacji. Zatem, nie przyjmując reprezentacji za oczywistość, perspektywa dynamiczna oferuje szersze teoretyczne pole działania. Z jednej strony daje możliwość zrozumienia, czym są reprezentacje oraz kiedy i jak powstają. Z drugiej strony może się okazać, że przynajmniej w niektórych przypadkach rolę odgrywanych przez stany wewnętrzne agenta dynamicznego po prostu nie można zinterpretować w użyteczny sposób jako reprezentacyjnych. Jaki jest związek pomiędzy podejściem do poznania opartym na SED a bardziej znanym podejściem obliczeniowym i koneksjonistycznym? Porównanie takie nastęrcza trudności. Na przykład musimy rozróżnić elementy matematyki leżące u podstaw każdego z tych podejść od twierdzeń teoretycznych, które te podejścia wysuwają. Podobnie jak formalizmy matematyczne, systemy obliczeniowe, koneksjonistyczne i dynamiczne mają mniej więcej równoważną moc w tym sensie, że każdy z nich

może zostać użyty do konstruowania modeli tej samej klasy zjawisk. Zatem nie ma użytecznego matematycznego rozróżnienia między tymi różnymi podejściami. Myślę, że jest to jeden z powodów, dla których dynamiczna hipoteza van Geldera zawodzi. Ponadto musimy przyznać, że obliczeniowość, koneksjonalizm i dynamika w ogóle nie są teoriami naukowymi, ponieważ same w sobie nie dostarczają przewidywań, które można łatwo falsyfikować. Są one raczej tym, co nazwałem ramami teoretycznymi (Beer 1995b). Dostarczają zestawu przedteoretycznych intuicji, słownictwa teoretycznego, stylu wyjaśniania, światopoglądu, w ramach którego formułowane i analizowane są poszczególne falsyfikowalne teorie konkretnych zjawisk poznawczych. Na przykład ramy obliczeniowe kładą nacisk na strukturę i treść wewnętrznych reprezentacji używanych przez agenta oraz algorytmy, za pomocą których manipuluje się tymi reprezentacjami. W przeciwieństwie do tego, struktura koneksjonistyczna kładzie nacisk na architekturę sieci, algorytm uczenia się, protokół szkoleniowy i opracowane pośrednie reprezentacje rozproszone. W tym sensie wiele modeli koneksjonizmu nadal ma charakter bezcielesny, niesytuowany i obliczeniowy (aczkolwiek rozproszony). Wreszcie ramy SED kładą nacisk na strukturę przestrzeni wszystkich możliwych trajektorii układu mózg – ciało – środowisko oraz różne siły, zarówno wewnętrzne, jak i zewnętrzne w stosunku do agenta, które kształtują te trajektorie w celu ustabilizowania określonego wzorca zachowania. Jest prawdopodobne, że wszystkie trzy perspektywy będą ważne w każdej przyszłej teorii zachowania i poznania. Na przykład, ponieważ neuronowe składniki modelu SED są często powtarzającymi się sieciami koneksjonistycznymi i ponieważ deliberatywne rozumowanie jest jednym ze zjawisk poznawczych, którym w końcu trzeba się zająć, idee i narzędzia matematyczne zarówno z koneksjonizmu, jak i obliczeniowości prawdopodobnie odegrają zasadniczą rolę nawet w teorii skupionej na SED. Dokładne połączenie spostrzeżeń z tych trzech ram teoretycznych (lub innych ram, których jeszcze nie można sobie wyobrazić!), które ostatecznie okażą się najbardziej owocne, pozostaje otwartą kwestią, którą mogą rozwiązać jedynie trwające badania empiryczne.

Zagadnienia metodologiczne

Poważne traktowanie ram SED rodzi wiele trudnych problemów metodologicznych. Badanie tylko jednego elementu układu mózg–ciało–środowisko jest wystarczająco trudne, ale jednoczesne badanie interakcji wszystkich trzech elementów jest zadaniem zniechęcającym. Eksperymentalnie brakuje nam obecnie instrumentów do monitorowania i manipulowania aktywnością wszystkich odpowiednich neuronów w układach nerwowych nienaruszonych, zachowujących się zwierząt, nie mówiąc już o odpowiednich właściwościach ciała i środowiska zwierzęcia. Teoretycznie obecnie brakuje nam narzędzi matematycznych niezbędnych do zrozumienia dużych sieci ściśle ze sobą powiązanych, heterogenicznych, nieliniowych elementów dynamicznych, szczególnie w systemach, które ewoluowały ze względu na ich skuteczność behawioralną, a nie ze względu na ich zrozumiałość w kategoriach tradycyjnych zasad projektowania inżynierskiego, takich jak modułowość i rozkład hierarchiczny. Z tych powodów wielu badaczy zwróciło się ku badaniu agentów modelowych przy użyciu dynamicznych sieci neuronowych i algorytmów ewolucyjnych. W tym podejściu modelowy „układ nerwowy” jest ucieleśniony w ciele modelowym, które z kolei jest usytuowane w modelowym środowisku. Cały system ewoluuje, aby wykonywać pewne interesujące zachowania. Powszechnie wybieranym modelem układu nerwowego są sieci neuronowe rekurencyjne w czasie ciągłym, o których wiadomo, że są uniwersalnymi aproksymatorami gładkiej dynamiki. Zazwyczaj ewoluują tylko parametry neuronowe, ale w niektórych pracach ewoluuje także architektura sieci i właściwości ciała. Istotną zaletą podejścia ewolucyjnego jest to, że minimalizuje ono założenia teoretyczne aprioryczne, a tym samym umożliwia badanie przestrzeni możliwych systemów środowiska mózg-ciało zdolnych do generowania określonego zachowania. Ta ewolucyjna metodologia została już z powodzeniem zastosowana do szerokiego zakresu interesujących zachowań. Wiele prac skupiało się na zachowaniach sensomotorycznych, takich jak orientacja, poruszanie się na nogach, unikanie obiektów i nawigacja.

Inny kierunek prac skupiał się na ewolucji zachowań związanych z uczeniem się. Ponadto poczyniono znaczne prace nad zachowaniami kierowanymi wzrokowo (Cliff i in. 1993) i ich zastosowaniem do percepcji kategorycznej, uwagi selektywnej i innych zadań interesujących poznawczo. Wreszcie, aktywnym obszarem badań jest także ewolucja komunikacji. Tak więc, chociaż istnieją trudne otwarte kwestie związane ze skalowaniem ewolucyjnych podejść do coraz bardziej skomplikowanych zachowań, można argumentować, że czynniki, które już wyewoluowały, są na tyle interesujące, że ich wnikliwa analiza może nauczyć nas wielu rzeczy na temat dynamiki układów mózg-ciało-środowisko. W istocie moim głównym zainteresowaniem nie jest ewolucja takich agentów modelowych per se, ale raczej analiza powstałych w ten sposób układów mózg-ciało-środowisko przy użyciu narzędzi teorii systemów dynamicznych. Podstawowym celem takiej analizy jest zbudowanie intuicji, koncepcji teoretycznych oraz narzędzi matematycznych i obliczeniowych niezbędnych do zrozumienia dynamiki układów mózg-ciało-środowisko. Chociaż czas letni zapewnia solidną podstawę dla takich dochodzeń, należy uwzględnić wiele dodatkowych kwestii. Na przykład istnieją różne poziomy, na których można analizować dynamikę układu mózg-ciało-środowisko, w tym autonomiczną dynamikę całego sprzężonego układu, w jaki sposób sprzężone zachowanie wynika z interakcji między nieautonomicznym środowiskiem a dynamiką agenta, w jaki sposób dynamika agenta nieautonomicznego wynika z interakcji między ciałem nieautonomicznym a dynamiką neuronową oraz w jaki sposób dynamika nieautonomicznego neuronu wynika z architektury, parametrów wewnętrznych i synaptycznych elementów neuronowych. Ostatnią kwestią, którą należy się zająć, jest zrozumienie dynamiki nieautonomicznej. Narzędzia matematyczne DST są najbardziej rozwinięte w przypadku autonomicznych układów dynamicznych, gdzie analiza może skupiać się na atraktorach i ich bifurkacjach. Jednakże, jak wspomniano powyżej, jeśli chcemy zrozumieć udział określonego składnika układu mózg-ciało-środowisko, musimy rozłożyć sprzężony system na oddziałujące na siebie nieautonomiczne podsystemy i zbadać ich przejściowe reakcje na zmieniające się w czasie sygnały wejściowe otrzymywane z inne komponenty. Niestety, matematyczne narzędzia do analizy dynamiki stanów przejściowych wymagają znacznego dalszego rozwoju.

Horyzont

Podobnie jak komputerowość i koneksjonizm, usytuowana, ucieleśniona i dynamiczna struktura opisana w tym rozdziale ma swoje korzenie w ideach, które po raz pierwszy sformułowano w latach czterdziestych i pięćdziesiątych XX wieku. Ponieważ jednak nowoczesna forma frameworku SED pojawiła się dopiero w latach 1985–1995, miała ona znacznie mniej czasu na rozwój niż ramy obliczeniowe i koneksjonistyczne. Liczba osób pracujących w ramach SED jest obecnie również znacznie mniejsza. Pomimo tych wad, idee umiejscowione, ucieleśnione i dynamiczne mają duży wpływ na myślenie w naukach kognitywnych, sztucznej inteligencji i robotyce, neuronauce, psychologii rozwojowej i filozofii umysłu. W celu dalszego zbadania zakresu i ograniczeń ram SED oraz wyjaśnienia najlepszego połączenia idei obliczeniowych, koneksjonistycznych i SED niezbędnych do zrozumienia mechanizmów zachowania i poznania, konieczny jest znaczny dalszy rozwój. Przede wszystkim będzie to wymagało zbudowania i analizy znacznie większej liczby konkretnych agentów modelowych, zwłaszcza tych o bardziej interesującym poznawczo charakterze. To z kolei będzie wymagało ciągłego rozwoju technik skalowania technik ewolucyjnych i analizy dynamicznej do większych systemów oraz dalszego rozwoju technik analizy przejściowej dynamiki nieautonomicznych układów dynamicznych. Wreszcie istnieje potrzeba lepszej edukacji w zakresie koncepcji systemów dynamicznych w społeczności nauk kognitywnych oraz oprogramowania wspierającego dynamiczną analizę systemów mózg-ciało-środowisko.