Przegląd Filozoficzny – Nowa Seria
R. 22: 2013, Nr 2 (86), ISSN 1230–1493
DOI: 10.2478/pfns-2013-0040

Marcin Miłkowski

Wyjaśnianie w kognitywistyce

Słowa kluczowe: *obliczanie, przetwarzanie informacji, mechanicyzm, wyjaśnianie obliczeniowe, rozwiązywanie problemów, biorobotyka*

Zamierzam bronić tezy, że podstawowym rodzajem wyjaśniania w kognitywistyce jest wyjaśnianie działania mechanizmów przetwarzania informacji. Mechanizmy te stanowią złożone, zorganizowane układy, których funkcjonowanie zależy od interakcji ich części i zachodzących w nich procesów.

Kognitywistyka (*cognitive science*) jest interdyscyplinarnym konglomeratem, obejmującym psychologię, lingwistykę, informatykę, robotykę, antropologię, filozofię, neuronauki, etologię, cybernetykę itd. Już wczesne opracowania na temat kognitywistyki, takie jak wpływowy raport dla Fundacji Sloana, ujmowały ją jako z zasady interdyscyplinarne badania nad poznaniem (Arbib i in. 1978). Przy całej wielości podejść kognitywistykę spaja przedmiot badań, czyli poznanie. Przedstawię argumentację, że właśnie dlatego jednym z charakterystycznych – lecz nie jedynym – rodzajem wyjaśniania w kognitywistyce jest wyjaśnianie procesów poznawczych w kategoriach mechanizmów przetwarzania informacji, czyli mechanizmów obliczeniowych.

Tekst ma następującą strukturę. W pierwszej części skrótowo opisuję wyjaśnienie zdolności poznawczej znane z klasycznych badań kognitywistycznych Allena Newella i Herberta Simona. Na jego przykładzie wykładam tezę bronionego przeze mnie podejścia do wyjaśniania, mianowicie podejścia mechanicystycznego. Pokazuję skrótowo, czym ta koncepcja różni się od tradycyjnego ujęcia wyjaśniania w kategoriach funkcjonalistycznych. Następnie wprowadzam drugi przykład, zaczerpnięty z aktualnych badań nad zdolnościami fonotaktycznymi świerszczy, które wyjaśnia się przy użyciu modeli robotycznych. Przykład ten posłuży mi do pokazania, że mechanicyzm akcentuje inne aspekty wyjaśniania w kognitywistyce niż funkcjonalizm, a także prowadzi – w odróżnieniu od funkcjonalizmu – do pluralizmu eksplanacyjnego.

1. Kryptoarytmetyka

Badając procesy poznawcze, Newell i Simon wybierali dobrze zdefiniowane problemy intelektualne. Celem ich badań było opracowanie teorii rozwiązywania problemów przez człowieka. Zaznaczali, że w ten sposób nie badają percepcji czy procesów sensomotorycznych, gdyż te mają inną specyfikę (Newell, Simon 1972).

Jednym z takich zadań jest tzw. kryptoarytmetyka. Zadanie kryptoarytmetyczne może mieć na przykład postać:

```
    SEND
+   MORE
=  MONEY
```

Za słowami kryje się operacja arytmetyczna na liczbach dziesiętnych. Zadanie polega na wskazaniu, jakie cyfry odpowiadają poszczególnym literom. Czytelników, którzy zaczęli się teraz zastanawiać nad rozwiązaniem, muszę przestrzec, że przeciętnemu badanemu dojście do poprawnej odpowiedzi zajmuje około pół godziny[1]. Newell i Simon zbadali, jak takie zagadki rozwiązują różni badani; okazało się, że stosują jedną z kilku możliwych reprezentacji zadania: jako proste podstawianie liter przez cyfry, jako równanie algebraiczne – i tak dalej. Hipotezą tych badaczy było, że rozwiązywanie problemów polega na poszukiwaniu rozwiązania w swoistej wielowymiarowej przestrzeni; kolejne operacje systemu przetwarzania informacji mają przybliżać do tego rozwiązania. Na podstawie uzyskanych w badaniach raportów słownych (a później – także danych dotyczących ruchu gałek ocznych) zbudowano program komputerowy, który miał symulować działanie jednego badanego. Nie uśredniano wyników całej grupy, gdyż istnieją duże różnice indywidualne, wyrażające się m.in. różnymi sposobami reprezentacji problemu.

Zbudowany program dobrze odzwierciedlał operacje opisane w raportach słownych, a jeszcze lepiej – jak się okazało – ruchy gałek ocznych (Newell, Simon 1972: 326–327), gdyż raporty słowne zwykle są niepełne (trudno jednocześnie myśleć i mówić). Co ciekawe, mimo że program odzwierciedlał postępowanie pojedynczej osoby, to po ustaleniu sposobu reprezentacji problemu można było przewidzieć przebieg jego rozwiązywania przez wielu badanych, jako parametr wskazując po prostu przestrzeń wyszukiwania rozwiązań.

Program miał formę tzw. systemu produkcji, czyli systemu zawierającego reguły w rodzaju „znak «stop» → zatrzymaj się", „znak nakazu skrętu w lewo → skręć w lewo". Aktualnie wykonywana instrukcja z następnika reguły jest

[1] 9567 + 1085 = 10652.

dopasowywana na podstawie bieżącego stanu systemu (np. obserwowanego znaku drogowego). Systemy takie służą do symulowania ludzkiego myślenia; a że hipotezą teoretyczną Newella i Simona było, że ludzie posługują się nie tyle idealnymi algorytmami do rozwiązywania problemów, ile zawodnymi i szybkimi heurystykami, które zwykle im wystarczają, to reguły w ich systemach były heurystyczne. Korzystając z takich właśnie heurystyk, opisywanych w postaci reguł, program dochodził do rozwiązania zadania kryptoarytmetycznego.

Podsumujmy najważniejsze cechy tego wyjaśnienia. Przeprowadzana jest analiza natury zadania; w tym celu bada się możliwe jego reprezentacje i poprawne sposoby jego rozwiązywania. Ważne jest też zebranie danych empirycznych na temat zachowania ludzi podczas rozwiązywania problemu; mają one tutaj zazwyczaj postać protokołów słownych (Ericsson, Simon 1995), niekiedy wspomaganych przez dane na temat ruchu gałek ocznych. Następnie budowany jest program komputerowy, którego poprawność jako wyjaśnienia bada się, testując, czy generuje prawdziwe przewidywania danych behawioralnych (tym razem takich, które nie służyły do generowania programu).

2. Mechanicyzm

Najgoręcej dyskutowanym w ostatnich latach ujęciem wyjaśniania w naukach szczegółowych jest ujęcie mechanistyczne (Machamer, Darden, Craver 2011; Bechtel 2008; Craver 2007; Glennan 2002). Co prawda zwolennicy mechanicyzmu (zwanego też niekiedy „neomechanicyzmem" w celu odróżnienia tej teorii od XVII-wiecznego mechanicyzmu) różnią się poglądami dosyć znacznie, lecz podstawowe zasady wyjaśniania mechanistycznego opisują stosunkowo podobnie. Jest to mianowicie rodzaj wyjaśniania przyczynowego, które dotyczy układów złożonych z wielu części, przejawiających jako całość pewną dyspozycję. Dyspozycja ta jest wyjaśniania przyczynowo jako rezultat oddziaływania między częściami mechanizmu, jego organizacji i zachodzących w nim procesów. Argumentuje się też, że wyjaśnianie takie jest typowe dla nauk szczegółowych, w których trudno dopatrywać się powszechnie obowiązujących praw przyrody, lecz gdzie często pojawiają się pewne generalizacje dotyczące mechanizmów odpowiedzialnych za różnego rodzaju zjawiska. Szczególnie wyrazistymi przykładami takich nauk mają być biologia, neurobiologia, neuropsychologia i psychologia; wskazuje się też często interdyscyplinarne badania kognitywistyczne, na których się tu skupię.

Swoistą odmianą wyjaśniania mechanistycznego jest wyjaśnianie zjawisk przetwarzania informacji przez odwołanie się do ich mechanizmów (Miłkowski 2013; Piccinini 2007). Przy wszelkich różnicach terminologicznych między moim podejściem a koncepcją Piccininiego ogólny schemat pozostaje ten sam: mechanizm przetwarzania informacji to układ złożony, którego struktura

przyczynowa odpowiada strukturze opisywanej za pomocą odpowiedniego modelu obliczeń. Członami relacji przyczynowych są bądź to stany układu, które odpowiadają stanowi maszyny abstrakcyjnej, bądź to inne cechy fizyczne, które można odnieść do abstrakcyjnego opisu modelu obliczeń (jeśli np. maszyna nie jest zdefiniowana w kategoriach przejść stanów). Istotne jest to, że stosowane w informatyce modele obliczeń zazwyczaj będą stanowić jedynie część właściwą pełnego modelu przyczynowego danego układu; oprócz tego w pełnym wyjaśnieniu będzie należało uwzględnić pewien schemat realizacji mechanizmu obliczeniowego, np. obejmujący rodzaj części elektronicznych czy typ procesora. Pełne wyjaśnienie mechanistyczne to kompletny opis struktury przyczynowej istotnej dla powstania badanego zjawiska. Tu warto zauważyć, że np. kolor klawiatury komputera nie będzie istotny dla wyjaśnienia, w jaki sposób wykonuje on program edytora tekstu, dlatego też nie będzie on uwzględniony; pełność opisu nie oznacza uwzględniania wszelkich dostępnych informacji, a jedynie informacji niezbędnych ze względu na wyjaśniane zjawisko.

Mechanicystyczne wyjaśnienie zjawiska polega bowiem nie tyle na podaniu warunków początkowych i uniwersalnego prawa przyrody, z których koniunkcji wywnioskować można opis wyjaśnianego zjawiska, ile na wskazaniu struktury przyczynowej danego procesu. Ponieważ obecnie najpełniej opracowaną koncepcją wyjaśniania przyczynowego jest interwencjonizm (Woodward 2003; Pearl 2000; Spirtes, Glymour, Scheines 2000), który jest w stanie poradzić sobie także z przytłaczającą większością trudności, z jakimi borykają się inne teorie przyczynowości (Halpern, Pearl 2005a, 2005b), mechanicyści chętnie sięgają po interwencjonistyczną eksplikację pojęcia przyczynowości. Eksplikacja jest tym bardziej przydatna, że jest opracowana również od strony matematycznej (istnieje nawet formalny język programowania, służący do wnioskowania o przyczynowości w takim ujęciu – por. Pearl 2000). Oczywiście pozostaje kwestią otwartą, czy spór między teorią dedukcyjno-nomologiczną wyjaśniania a mechanicyzmem nie jest nieco przerysowany, gdyż – przy odpowiednim osłabieniu pojęcia prawa, na co nie zgodzą się zapewne niektórzy zwolennicy uniwersalności praw (Woodward 2002) – ujęcia te stają się komplementarne (Andersen 2011). Można bowiem sądzić, że przyczynowość zachodzi ze względu na pewne prawidłowości, choćby nie były one powszechne i ujawniały się jedynie w pewnych swoistych układach, ograniczonych w czasie i przestrzeni (co ogranicza zarazem zasięg zastosowania opisu tych prawidłowości w wyjaśnianiu i predykcji).

Wróćmy do mechanizmów przetwarzania informacji. I tak, zdolność ludzi do rozwiązywania zadania typu SEND + MORE = MONEY opiera się na zdolności do przetwarzania informacji o literach i cyfrach. System przetwarzania informacji opisany przez Newella i Simona ma odpowiadać dostępnym

na ogół w świadomości sekwencyjnym i szeregowym procesom poszukiwania rozwiązania. Jest to więc opis mechanizmu poznawczego u badanych, którzy sobie z tym zadaniem radzą. Analiza zadania doprowadza do powstania opisu wyjaśnianej zdolności poznawczej. Tradycyjnie taką zanalizowaną zdolność poznawczą określano w kognitywistyce mianem „kompetencji". Uzyskane dane na temat zachowania (m.in. z raportów słownych) służą do stworzenia listy reguł produkcji stosowanych w programie wyjaśniającym tę kompetencję; są także wykorzystywane do konfirmacji teorii realizacji zadania.

Z punktu widzenia mechanicyzmu wyjaśnienie proponowane przez Newella i Simona nie jest kompletne. Nie dlatego, że ich badania należą do tzw. nurtu symbolicznego w kognitywistyce, który nie opisuje bezpośrednio procesów w układzie nerwowym, lecz postuluje istnienie poziomu opisu, na którym obowiązują psychologiczne prawidłowości wyższego rzędu. *A priori* oczywiście istnienia takiego poziomu wykluczyć nie sposób; sęk w tym, że po prostu jest to jedynie hipoteza stosowana w badaniach; potwierdza się ona tylko w ten sposób, że model kryptoarytmetyki skutecznie przewiduje i wyjaśnia ludzkie zachowanie. Dziś istnienie takiego poziomu budzi większe wątpliwości, lecz sceptyk musi odpowiedzieć na pytanie, dlaczego zatem model Newella i Simona przewiduje zachowanie.

Sceptyk, gdyby był mechanicystą, mógłby wskazać na pogwałcenie kilku norm wyjaśniania. Po pierwsze, może twierdzić, że dane eksperymentalne Newella i Simona pokazują jedynie, iż ich program komputerowy *wystarcza* do rozwiązania tego zadania. Mój kalkulator może zupełnie inaczej liczyć 4 + 4 = 8 niż człowiek, uzyskując przecież ten sam wynik. Ten zarzut jest jednak tylko o tyle słuszny, o ile ma dotyczyć umysłowych operacji elementarnych: Newell i Simon rozczłonkowują raport słowny na wiele części, z których każdą odnoszą do pewnej operacji elementarnej systemu produkcji. Innymi słowy, stosowana jest tutaj strategia dekompozycji złożonego procesu na operacje składowe: działanie programu komputerowego nie odpowiada ludzkiemu zachowaniu jedynie pod względem rezultatu (tak jak w przypadku działania kalkulatora liczącego sumę dwóch liczb), lecz również pod względem przebiegu kolejnych operacji składowych. Model Newella i Simona ma zatem być nie tylko słabo równoważny zachowaniu poznawczemu, mając te same wejścia i wyjścia (Fodor 1968), lecz również mocno równoważny całemu procesowi. Ta równoważność jednak nie jest całkowicie dowiedziona, gdyż nie wiemy, dlaczego te, a nie inne operacje składowe są elementarne.

Tu sceptyk-mechanicysta mógłby wskazywać, że potrzebne wyjaśnienie mechanistyczne jest wielopoziomowe; otóż zdolność wysokiego poziomu powinna zostać wyjaśniona konstytutywnie. Dlaczego system poznawczy człowieka ma te operacje jako elementarne? Gdybyśmy wyjaśniali działanie kalkulatora, odwołalibyśmy się do schematu ideowego i inżynierskiego (a nie

tylko fizycznego) opisu jego części elektronicznych. Newell i Simon starają się ograniczyć arbitralność doboru operacji elementarnych, stosując m.in. uzyskane w badaniach wyniki w postaci ograniczonej pojemności pamięci krótkoterminowej (Miller 1956) czy też liczby operacji wysokopoziomowych wykonywanych przez system poznawczy na sekundę (Newell, Simon 1972: 808–809). To może ich wyjaśnienie uwiarygodnić, lecz nie stanowi jego rzeczywistej konfirmacji. Tymczasem mechanicyzm wymaga, aby istniały empiryczne świadectwa potwierdzające, że konstytutywny poziom mechanizmu – odpowiadający za elementarne operacje przetwarzania informacji w mechanizmie – jest taki, a nie inny. W przeciwnym razie jest to jedynie wyjaśnienie wiarygodne, a nie rzeczywiste.

Tu mechanicyzm dosyć wyraźnie różni się od funkcjonalistycznych ujęć wyjaśniania (Piccinini, Craver 2011). Większość funkcjonalistów sądziła (i sądzi), że to, jak realizowany jest proces przetwarzania informacji, nie jest istotne dla jego wyjaśnienia, gdyż procesy te są wielorako realizowalne; stąd też autonomia psychologii i kognitywistyki (Fodor 2008). Uważano, że należy jedynie przedstawić analizę czy też opis funkcjonalny, który wystarcza do ujawnienia się kompetencji poznawczej. Ten funkcjonalizm ma jednak bardzo niefortunne konsekwencje; do zabicia prezydenta Kennedy'ego wystarczy bomba atomowa, lecz opis wybuchu bomby atomowej nie ma żadnej mocy eksplanacyjnej w odniesieniu do faktycznego zamachu. Podobnie sposób działania liczydła nie tłumaczy zdolności arytmetycznych u dzieci: ich systemy poznawcze (o ile mi wiadomo) nie mają przecież nic, co odpowiadałoby wprost kilku szeregom drewnianych kulek. Innymi słowy, tu funkcjonalista raczej nie obroni Newella i Simona; może tylko zauważyć, że uzyskanie niezbędnych danych empirycznych było po prostu w latach 70. niemożliwe. Dlatego i dziś wiele wyjaśnień jest jedynie wiarygodnych, a nie rzeczywistych (jeśli w ogóle istnieją wyjaśnienia, które tę normę mechanicyzmu spełniają).

Mówiąc inaczej, mechanicyści będą najczęściej wymagać integracji wyjaśnień kognitywistycznych z neuropsychologicznymi, przez co nie będzie mowy o arbitralności doboru opisów mechanizmu – na poziomie jego składników czy też jego funkcjonowania w otoczeniu – jedynie wystarczających do trafnego opisania wyjaśnianego zjawiska. Nie chodzi tu o zwykłe odrzucenie tezy o wielorakiej realizacji, choć jest to charakterystyczne dla wielu, choć nie wszystkich mechanicystów; przeciwnikami obrony tezy o wielorakiej realizacji z pozycji mechanicystycznych są Bechtel i Mundale (1999), ale wśród mechanicystów istnieją też zwolennicy tej tezy (Wilson, Craver 2007; Gillett 2003). Czy przyjmujemy tezę o wielorakiej realizacji, czy nie, w wyjaśnieniu konstytutywnym musimy opisać, jak dany mechanizm jest konstytuowany: sama wieloraka realizacja nie ma tu wiele do rzeczy; jeśli bowiem zachodzi,

to istnieje coś, co łączy różne realizacje, i dzięki czemu są one realizacjami tej samej zdolności poznawczej. Utrzymując, że poziom konstytutywny nie może zostać pominięty w wyjaśnianiu, nie wymagamy jednak koniecznie, żeby był to poziom neurobiologiczny; gdyby się okazało, że skądinąd wiadomo, jakie umysłowe operacje elementarne istnieją, schodzenie na poziom neurobiologiczny byłoby fanaberią, tak samo jak niepotrzebną fanaberią w wyjaśnianiu funkcjonowania pompki do roweru byłoby odwoływanie się do równań mechaniki kwantowej. Poziom najniższy danego wyjaśnienia jest ustalany w praktyce naukowej, nie zaś dany z góry (Machamer, Darden, Craver 2011); ten rodzaj integracji wielu poziomów organizacji nie pociąga za sobą wiary w tezę o istnieniu jednego fundamentalnego poziomu opisu rzeczywistości. Z tego też względu mechanicyzm nie dezawuuje *a priori* wyjaśnień w stylu Newella i Simona czy też innych wczesnych prac kognitywistycznych (Miller, Galanter, Pribram 1980), w których postulowano istnienie bazowego poziomu psychologicznego, odrębnego jednak od poziomu neurobiologicznego. Rzecz w tym, że wiarygodność hipotezy o istnieniu takiego poziomu dziś jest niższa niż w latach sześćdziesiątych czy siedemdziesiątych ubiegłego wieku.

Tak ujmując sprawę, mechanicyzm nie ma problemu typowego dla funkcjonalizmu, który stoi przed dylematem: albo będzie się bronić autonomii nauk szczegółowych opisujących funkcje, ale za cenę akceptacji możliwych epifenomenalnych czynników w wyjaśnianiu (takich jak bomba w wyjaśnieniu zamachu na Kennedy'ego), albo trzeba będzie odrzucić funkcjonalizm na rzecz teorii identyczności (Hensel 2011). Co gorsza, antyredukcjonizm może też prowadzić do zarzutu możliwej banalizacji każdego opisu w kategoriach funkcjonalnych (Godfrey-Smith 2008): jak pokazał Putnam, jeśli nie ma żadnych ograniczeń w tworzeniu opisów funkcjonalistycznych, a ze względu na programowy antyredukcjonizm żadnych ograniczeń się nie przyjmuje, wszystko może okazać się takim lub owym układem funkcjonalnym, np. komputerem – istnieje dosyć prosty dowód, że dowolny otwarty układ fizyczny można opisać jako automat o skończonej liczbie stanów (Putnam 1991). Banalizacji można uniknąć dopiero, gdy przyjmie się ograniczenie, że dobre są tylko te opisy funkcjonalne, które zawierają istotne przyczynowo czynniki, konstytuowane na poziomie niższym. Wówczas dowodu Putnama nie da się przeprowadzić: postuluje on byty, których opis narusza po prostu metodologię wyjaśnień przyczynowych i których nie sposób wskazywać na poziomie niższym bez wskazywania na zasady opisu z poziomu wyższego (Miłkowski 2013: rozdz. 2). Z tego też względu mechanicyzmowi blisko do pewnej postaci teorii identyczności typów, choćby była ona traktowana raczej jako heurystyka niż prawda aprioryczna (Bechtel, McCauley 1999; McCauley, Bechtel 2001). Innymi słowy, założenie o identyczności (zwłaszcza typów) możemy uchylić, ale służy nam ono jako przewodnik w odkrywaniu mechanizmów. Bez jakiej-

kolwiek identyczności nie ma jednak mechanizmu, a zatem i wyjaśnienia nie są warte funta kłaków.

Kolejną normą naruszoną przez Newella i Simona jest zasada wymagająca, aby wyjaśnianie nie zawierało luk ani terminów-wytrychów w rodzaju „reprezentacja" (bliżej niesprecyzowana), „aktywacja" czy „komunikacja". Otóż operacje czytania i rozumienia napisów są traktowane przez Newella i Simona jako pierwotne; są one zatem desygnowane takimi terminami-wytrychami. Mogą zostać sprecyzowane dopiero w kolejnych badaniach, które wyjaśnią, w jaki sposób człowiek czyta cyfry i litery. Zauważmy, że takie terminy-wytrychy pojawiają się zwykle w pierwszych modelach procesów poznawczych i są potem stopniowo eliminowane. Z tego też względu opis mechanizmu rozwiązywania problemów u Newella i Simona nie jest kompletny, lecz schematyczny; ze względu na brak opisu poziomu konstytutywnego jest wręcz szkicowy (brak danych o tym, co w istocie odpowiada za operacje elementarne).

Podsumujmy zasady mechanicyzmu. Wyjaśnianie działania mechanizmu musi obejmować zarówno odniesienie do środowiska, w którym mechanizm występuje, jak i roli, jaką w nim odgrywa. To odniesienie powinno być uwzględnione w analizie zadania, którą przeprowadzali Newell i Simon; jest oczywiste, że zadanie jest poniekąd sztuczne, bo kryptoarytmetyka służy jedynie rozrywce i nie jest niezbędna do życia. Dlatego nie ma ona żadnej roli adaptacyjnej ani szczególnego znaczenia środowiskowego. Ta kompetencja po prostu nieszczególnie do czegokolwiek się przydaje. Aby w pełni wyjaśnić, jak ta kompetencja się pojawia, należy wyjaśnić z kolei procesy przetwarzania informacji zachodzące wewnątrz samego mechanizmu osadzonego w środowisku. Zazwyczaj wyjaśnienie na tym poziomie ma postać modelu obliczeniowego, na przykład programu komputerowego (w klasycznych podejściach) lub wytrenowanej sieci neuropodobnej. Jednak na tym poziomie wyjaśnienie się nie kończy. Do wyjaśnienia pozostaje, jak realizowany jest sam program (lub jakie procesy odpowiadają za przetwarzanie informacji w sieci neuronalnej). U Newella i Simona mamy tylko niewyraźne zarysy tego poziomu realizacyjnego. Inna rzecz, że Newell jest tego świadomy i za palący problem uważa wskazanie, jak układ nerwowy staje się komputerem symbolicznym (Newell 1980).

3. Fonotaksja

Jak widać, mechanicyzm nie jest szczególnie miłościwy dla tradycyjnych modeli kognitywistycznych, mimo że uznaje, iż mogą one dostarczyć wyjaśnień wiarygodnych, choćby i jedynie szkicowych. Przyjrzę się teraz wyjaśnieniu fonotaksji u świerszczy, odwołującemu się do modelu robotycznego (Webb 1995). Biorobotyka, czyli wyjaśnianie zdolności poznawczych przez

budowanie robotów symulujących systemy biologiczne, jest dziedziną stosunkowo młodą (Webb 2000, 2002), lecz jej korzenie sięgają behawiorystycznych sztucznych modeli Tolmana (1939). Symulacje bywają rzeczywiste, kiedy buduje się fizyczny model, a także komputerowe, czyli takie, które nie wymagają istnienia robota w rzeczywistości. Te pierwsze mają tę przewagę, że budowanie robota jest jednocześnie eksperymentem, który pokazuje, że posiadany model robota jest kompletny, bowiem wystarcza do jego skonstruowania (inna sprawa, że w trakcie budowy czyni się wiele założeń *ad hoc*, dzięki którym on działa, ale które niekoniecznie mają jakikolwiek związek z modelowanym organizmem).

Webb wyjaśnia zdolność samic świerszczy do fonotaksji, czyli poruszania się w kierunku źródła dźwięku wydawanego przez samca. Okazuje się, że istotna w tym wypadku jest nie tylko charakterystyka samego dźwięku (częstotliwość 4–5 kHz), ale także budowa narządu słuchu u świerszcza. Świerszcze mają dwie pary uszu, z których jedna umiejscowiona jest na odwłoku w okolicach przetchlinek, a druga na przednich odnóżach. Są one połączone ze sobą, co zapewnia im dużą kierunkowość; ich fizyczna budowa jest typowa dla doskonale znanego inżynierom różnicowego czujnika ciśnienia, co ułatwiło replikację tego układu w robocie. Taka budowa narządu słuchu jest zresztą dosyć powszechna w świecie organizmów żywych (Michelsen, Larsen 2008). Wiadomo też, jaka jest budowa układu nerwowego świerszcza i jego narządów motorycznych.

Wszystkie te dane zostały wykorzystane przy konstrukcji robota, którego pierwsza wersja była dość mocno wyidealizowana – na przykład wykorzystywano napęd na koła. Zachowaniem robota rządził początkowo bardzo uproszczony układ: lewy neuron słuchowy pobudzał lewy neuron motoryczny i powodował inhibicję prawego – i na odwrót. To daleko idąca idealizacja, bo użycie synaps pobudzających i hamujących z jednego neuronu nie jest spójne z danymi biologicznymi (Webb 2008: 28). W kolejnych modelach idealizacje te były stopniowo zmniejszane. Za pomocą sztucznego układu udało się jednak powtórzyć wyniki wielu eksperymentów z żywymi świerszczami; robot poruszał się w stronę żywego samca tak jak biologiczna samica.

Warto zauważyć, że przetwarzanie informacji docierających z narządu słuchu nie jest jedynym składnikiem wyjaśnienia zachowania świerszcza. Równie istotne są fizyczne własności tego narządu, ośrodek rozpraszania fal dźwiękowych (pod wodą dźwięk rozchodzi się inaczej), a także wiedza na temat układu nerwowego owada. Co istotne, gdyby analizować zdolność świerszcza w oderwaniu od rzeczywistości biologicznej, można byłoby błędnie zakładać, że zadanie jest znacznie trudniejsze: owad musi bowiem jakoś odfiltrować nieistotne dźwięki z otoczenia i nie reagować na wszystkie bodźce tak samo. Filtrowanie jednak odbywa się w pewnej mierze fizycznie, dzięki budowie

ucha i jego wysokiej kierunkowości, przez co zadanie jest znacznie łatwiejsze i nie wymaga ogromnych mocy obliczeniowych, których świerszcz po prostu nie ma.

Rzecz jasna zjawisko fonotaksji jest bardzo proste w porównaniu z wieloma zdolnościami poznawczymi bardziej złożonych organizmów. To właśnie sprawia, że łatwo je replikować przy użyciu robota; ba, ograniczenia materiałowe sprawiają, że nie sposób wiernie odtworzyć też wielu prostych układów biologicznych. Można by co prawda mieć wątpliwości, czy fonotaksja w ogóle ma cokolwiek wspólnego z poznaniem. Tu chyba nie warto kruszyć kopii; przykład ten jest powszechnie przytaczany w podręcznikach kognitywistyki jako ilustracja idei poznania ucieleśnionego (Clark 2001: 103–108; Bermúdez 2010: 437–438). Nawet jeśli uznamy, że poznanie wymaga istnienia reprezentacji, a o reprezentacjach u świerszcza w tym wypadku raczej mowy nie ma (Miłkowski 2013: rozdz. 4), to jest to z pewnością przypadek przetwarzania informacji[2].

4. Informacja a pluralizm eksplanacyjny

Przy mechanistycznym ujęciu wyjaśniania procesów poznawczych staje się jasne, że wymaga ono zastosowania kilku różnych metod badawczych – do opisu przynajmniej trzech wspomnianych wyżej poziomów mechanizmów: kontekstowego, właściwego i konstytutywnego. Nic dziwnego więc, że kognitywistyka pozostaje interdyscyplinarnym konglomeratem wielu różnych dziedzin. Ten pluralizm wymusza sama natura wyjaśnienia.

Można mieć jednak w tym miejscu dwie wątpliwości. Czy uprzywilejowane miejsce pojęcia „przetwarzanie informacji" nie podważa pluralizmu eksplanacyjnego? I czy wymóg integracji wyjaśnień różnych poziomów nie prowadzi *de facto* do unifikacji, a unifikacja – do podważenia pluralizmu?

Odpowiem na te zarzuty po kolei. Zacznijmy od pojęcia informacji. Posługiwałem się w tym artykule pojęciem „informacja" w sposób techniczny, ale bez definicji; pora wreszcie na eksplikację, co miałem na myśli. Nie chodzi o informacje w sensie semantycznym (który pozostaje trudno uchwytny mimo istnienia już wielu teorii informacji semantycznych), lecz raczej strukturalnym. Otóż o informacji w sensie strukturalnym możemy mówić wtedy, gdy pewien fizyczny nośnik ma co najmniej dwa stany rozróżniane przez jakiś proces, lub też, co jest równoważne, że pewien proces reaguje na co najmniej dwa stopnie

[2] Owady mogą mieć jednak dość złożone struktury poznawcze i tzw. modele wyprzedzające (*forward models*), przez co część ich stanów będzie spełniać nawet wyśrubowane definicje pojęcia reprezentacji (o ile tylko, oczywiście, warunkiem posiadania reprezentacji nie będzie posługiwanie się językiem człowieka). Por. Webb 2004, 2006.

swobody danego nośnika. Tak strukturalną zawartość informacyjną definiuje MacKay (1969), który pokazuje, że ilekroć mowa o innych pojęciach informacji w teoriach sformalizowanych – w teorii Shannona czy Fishera – w istocie chodzi o to samo pojęcie, lecz w grę wchodzą różne *miary* informacji. Definiując przetwarzanie informacji, nie muszę posługiwać się żadną miarą; wymagam jedynie, aby mechanizm przetwarzający informacje po prostu reagował na pewien stan nośnika; stanów tych musi być więcej niż jeden. To znaczy, że przetwarzanie informacji nie wymaga istnienia ani żadnych statycznych nośników – informacje mogą nieść dziurki w kartach perforowanych, namagnesowane taśmy, rowki na płycie gramofonowej, poziomy napięcia, natężenia światła czy też impulsy generowane przez neurony. Mogą one być cyfrowe i analogowe; nie wykluczam bynajmniej istnienia informacji analogowych.

Takie pojęcie informacji jest bardzo szerokie i oczywiście można opisać dowolny proces jako informacyjny; jeśli istnieje jakaś obawa co do adekwatności tego pojęcia, to raczej będzie dotyczyć nadmiernego liberalizmu w jego użyciu niż jego zbyt wąskiego zakresu. Wbrew pozorom jednak duży zakres nie stwarza żadnych problemów z banalizacją pojęcia – informacja jest tu traktowana jak typowe pojęcie sformalizowane czy matematyczne, które można odnieść do czegokolwiek; np. liczyć możemy wszystko. Dlatego też uznaję, że jest to uprawniony sposób posługiwania się takim pojęciem.

Warto jednak zauważyć, że nie każdy mechanizm będzie mechanizmem przetwarzania informacji. Tylko mechanizm, którego zdolność na poziomie kontekstowym (w otoczeniu) można opisać w kategoriach niezależnych od fizycznego nośnika, będzie mechanizmem przetwarzania informacji. I tak wiadro z wodą i płynem do mycia podłogi w kuchni nie jest mechanizmem przetwarzania informacji, choć oczywiście można stworzyć nawet skomplikowany komputer, korzystając z rozwiązań hydraulicznych (Bissell 2007). Odkurzacz nie służy do tworzenia zbioru liczb losowych przez rotowanie cząsteczek kurzu w zbiorniku, bo stan tych cząstek nie wpływa regularnie na żaden inny proces; postulowanie takiego procesu nie ma sensu. Natomiast i płyta CD, i karta pamięci mogą być opisane niezależnie od nośnika jako przechowujące informacje. Ich struktura fizyczna może bowiem posłużyć pewnemu procesowi w odpowiedni sposób. Podobnie jest z całymi mechanizmami przetwarzania informacji; można zastępować podukłady złożone z lamp układami tranzystorowymi, o ile tylko wejścia i wyjścia układów zostaną odpowiednio dostosowane.

Mówiąc o przetwarzaniu informacji przez układ nerwowy, postulujemy istnienie takiego opisu jego działania, na którym nie musimy wchodzić w naturę nośnika, by powiedzieć, jak działa informacyjnie. Oczywiście natura nośnika jest też bardzo ważna (to poziom konstytutywny w neuronauce obliczeniowej), ale sam opis obliczeń jest tylko wtedy nienaciągany, gdy nie wszystkie wła-

sności poziomu niższego są uwzględniane w opisie systemu obliczeniowego, a jedynie te, które są kluczowe dla przetwarzania informacji.

Zauważmy, że mechanicyzm wcale nie wyklucza się tu z dynamicyzmem, a wręcz może niekiedy mu przyklasnąć. W świetle metodologii mechanicystycznej dobre jest dynamistyczne wyjaśnienie tzw. błędu „*A*-nie-B", do którego wyjaśnienia Piaget postulował poziom reprezentacji typu logicznego. Okazało się ono jednak lepiej wyjaśnialne w kategoriach przetwarzania sensomotorycznego (Thelen i in. 2001). W tym nowszym wyjaśnieniu nadal postuluje się przetwarzanie informacji, lecz innego rodzaju.

Przejdźmy do sprawy drugiej, czyli (pozornego) napięcia między pluralizmem a integracją czy unifikacją. Mechanicyści rzeczywiście podkreślają potrzebę integracji różnych dziedzin nauki i wskazują na różne heurystyki redukcjonistyczne, które pozwalają taką częściową integrację uzyskać (Bechtel, Richardson 1993). Dekompozycja i lokalizacja to podstawowe takie strategie. Jednak integracja między naukami nie ma służyć zastąpieniu ich jedną, większą nauką.

Po pierwsze, teorie mogą mieć różne cele i różnie idealizować swoje obiekty teoretyczne. Mogą być częściowo ze sobą zgodne i móc się wzajemnie ograniczać, lecz to nie znaczy, że będą wzajemnie sprowadzalne. Może nie jest to tak mocna teza, jak twierdzenie, że lis w ekologii jest zupełnie innym obiektem niż lis w biologii molekularnej (Paprzycka 2005), bo identyfikacja tego samego obiektu w różnych teoriach powinna być możliwa; lecz jedna teoria nie ma wcale wyprzeć drugiej, bo inne są jej cele poznawcze, co wpływa też na to, że mówi o innych aspektach lisa.

Po drugie, jeśli są to teorie logicznie i statystycznie niezależne, to pozbywając się jednej z nich, po prostu tracimy źródło informacji, które mogłoby służyć do niezależnego potwierdzania hipotez. Wydaje się, że w naukach szczegółowych dopiero konfrontując wiele źródeł informacji uzyskujemy mocne hipotezy (Wimsatt 2007). Irracjonalne byłoby zastępowanie wielu źródeł informacji jednym, jeśli są one od siebie niezależne.

Mówiąc krótko, redukcjonizm mechanicystyczny nie jest redukcjonizmem skrajnym; niekiedy wręcz potrafi występować w szatach antyredukcjonizmu (Craver 2005). Jest tak dlatego, że nie ma jednak zbyt wiele wspólnego z tradycyjną unifikacją, z przyszłą kompletną fizyką, lecz więcej z faktycznymi wysiłkami na rzecz zderzania różnych dziedzin opisujących te same lub podobne zjawiska w różny sposób i przy użyciu różnych metod eksperymentalnych.

Kognitywistyka, będąc konglomeratem wielu nauk mówiących o procesach przetwarzania informacji w różny sposób – informatyczny, cybernetyczny (w kategoriach teorii sterowania), neurobiologiczny, psychologiczny czy społeczny itd. – nie musi zostać sprowadzona do jednego mianownika. Wręcz nie powinna, jeśli nie chcemy rezygnować z różnych i niezależnych metod

empirycznych. Należy się raczej starać o pogłębianie teorii łączenia różnych, niekiedy sprzecznych ze sobą modeli w jednolitych ramach. Mechanicyzm to koncepcja, której właśnie taka wizja kognitywistyki odpowiada.

Bibliografia

Andersen Holly (2011), *The Case for Regularity in Mechanistic Causal Explanation*, „Synthese" 189, 3, s. 415–432, DOI: 10.1007/s11229-011-9965-x.

Arbib Michael, Baker Carl Lee, Bresnan Joan, D'Andrade Roy G., Kaplan Ronald, Keyser Samuel Jay, Norman Donald A. i in. (1978), „Cognitive Science".

Bechtel William (2008), *Mental Mechanisms*, New York, Routledge: Taylor & Francis Group.

Bechtel William, McCauley Robert N. (1999), *Heuristic Identity Theory (Or Back to the Future): The Mind-Body Problem against the Background of Research Strategies in Cognitive Neuroscience*, „Proceedings of the 21st Annual Meeting of the Cognitive Science Society", s. 67–72, Mahwah, NJ: Erlbaum.

Bechtel William, Mundale Jennifer (1999), *Multiple Realizability Revisited: Linking Cognitive and Neural States*, „Philosophy of Science" 66, 2, s. 175–207.

Bechtel William, Richardson R.C. (1993), *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton: Princeton University Press.

Bermúdez José Luis (2010), *Cognitive Science: An Introduction to the Science of the Mind*, Cambridge, New York: Cambridge University Press.

Bissell C. (2007), *The Moniac. A Hydromechanical Analog Computer of the 1950s*, „Control Systems Magazine. IEEE" 27, 1, s. 69–74.

Clark A. (2001), *Mindware: An Introduction to the Philosophy of Cognitive Science*, Oxford: Oxford University Press, USA.

Craver Carl F. (2005), *Beyond Reduction: Mechanisms, Multifield Integration and the Unity of Neuroscience*, „Studies in History and Philosophy of Science. Part C: Studies in History and Philosophy of Biological and Biomedical Sciences" 36, 2, s. 373–395, DOI: 10.1016/j.shpsc.2005.03.008.

Craver Carl F. (2007), *Explaining the Brain. Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Oxford University Press.

Ericsson Anders K., Simon Herbert A. (1995), *Analiza protokołów. Wprowadzenie i podsumowanie*, w: *Czy powrót introspekcji? Zbieranie i analiza danych słownych*, red. Tadeusz Tyszka, przeł. Józef Radzicki, Warszawa: Wydawnictwo Naukowe PWN, s. 116–175.

Fodor Jerry A. (1968), *Psychological Explanation: An Introduction to the Philosophy of Psychology*, New York: Random House.

Fodor Jerry A. (2008), *Nauki szczegółowe (albo: niejednorodność nauki jako hipoteza robocza)*, przeł. Marcin Gokieli, w: *Analityczna metafizyka umysłu*, red. Marcin Miłkowski, Robert Poczobut, Warszawa: Wydawnictwo IFiS PAN, s. 56–75.

Gillett C. (2003), *The Metaphysics of Realization, Multiple Realizability, and the Special Sciences*, „The Journal of Philosophy" 100, 11, s. 591–603, DOI: 10.1111/j.1747-9991.2007.00062.x.

Glennan Stuart (2002), *Rethinking Mechanistic Explanation*, „Philosophy of Science" 69 (S3), S342–S353, DOI: 10.1086/341857.

Godfrey-Smith Peter (2008), *Triviality Arguments against Functionalism*, „Philosophical Studies" 145, 2, s. 273–295, DOI: 10.1007/s11098-008-9231-3.

Halpern Joseph Y., Pearl Judea (2005a), *Causes and Explanations: A Structural-Model Approach. Part I: Causes*, „The British Journal for the Philosophy of Science" 56, 4, s. 843–887, DOI: 10.1093/bjps/axi147.

Halpern Joseph Y., Pearl Judea (2005b), *Causes and Explanations: A Structural-Model Approach. Part II: Explanations*, „The British Journal for the Philosophy of Science" 56, 4, s. 889–911, DOI: 10.1093/bjps/axi148.

Hensel Witold M. (2011), *Dwa funkcjonalizmy Hilary'ego Putnama, czyli kawałek historii z morałem*, „Diametros" 29, s. 31–49.

Machamer Peter, Darden Lindley, Craver Carl F. (2011), *Myślenie w kategoriach mechanizmów*, przeł. Witold M. Hensel, „Przegląd Filozoficzno-Literacki" 2–3 (31), s. 145–173.

MacKay Donald MacCrimmon (1969), *Information, Mechanism and Meaning*, Cambridge: MIT Press.

McCauley R.N., Bechtel W. (2001), *Explanatory Pluralism and Heuristic Identity Theory*, „Theory & Psychology" 11, 6, s. 736–760, DOI: 10.1177/0959354301116002.

Michelsen Axel, Larsen Ole Naesbye (2008), *Pressure Difference Receiving Ears,* „Bioinspiration & Biomimetics" 3: 011001, DOI: 10.1088/1748-3182/3/1/011001.

Miller George A. (1956), *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information*, „Psychological Review" 63, 2, s. 81–97, DOI: 10.1037/h0043158.

Miller George A., Galanter Eugene, Pribram Karl H. (1980), *Plany i struktura zachowania*, przeł. Aldona Grzybowska, Adam Szewczyk, Warszawa: Państwowe Wydawnictwo Naukowe.

Miłkowski Marcin (2013), *Explaning the Computational Mind*, Cambridge, Mass.: MIT Press.

Newell Allen (1980), *Physical Symbol Systems*, „Cognitive Science: A Multidisciplinary Journal" 4, 2, s. 135–183, DOI: 10.1207/s15516709cog0402_2.

Newell Allen, Simon Herbert A. (1972), *Human Problem Solving*, Englewood Cliffs, NJ: Prentice Hall.

Paprzycka Katarzyna (2005), *O możliwości antyredukcjonizmu*, Warszawa: Semper.

Pearl Judea (2000), *Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.

Piccinini Gualtiero (2007), *Computing Mechanisms*, „Philosophy of Science" 74, 4, s. 501–526, DOI: 10.1086/522851.

Piccinini Gualtiero, Craver Carl (2011), *Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches*, „Synthese" 183, 3, s. 283–311, DOI: 10.1007/s11229-011-9898-4.

Putnam Hilary (1991), *Representation and Reality*, Cambridge, Mass.: The MIT Press.

Spirtes Peter, Glymour Clark N., Scheines Richard (2000), *Causation, Prediction, and Search*, 2nd ed., Cambridge, Mass.: The MIT Press.

Thelen Esther, Schöner Gregor, Scheier Christian, Smith Linda B. (2001), *The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching*, „Behavioral and Brain Sciences" 24, 1, s. 1–34, DOI: 10.1017/S0140525X01003910.

Tolman E.C. (1939), *Prediction of Vicarious Trial and Error by Means of the Schematic Sowbug*, „Psychological Review" 46, 4, s. 318–336, DOI: 10.1037/h0057054.

Webb Barbara (1995), *Using Robots to Model Animals: A Cricket Test*, „Robotics and Autonomous Systems" 16, 2–4, s. 117–134, DOI: 10.1016/0921-8890(95)00044-5.

Webb Barbara (2000), *What Does Robotics Offer Animal Behaviour?* „Animal Behaviour" 60, 5, s. 545–558, DOI: 10.1006/anbe.2000.1514.

Webb Barbara (2002), *Can Robots Make Good Models of Biological Behaviour?* „Behavioral and Brain Sciences" 24, 6, s. 1033–1050; Discussion: 1050–1094, DOI: 10.1017/S0140525X01000127.

Webb Barbara (2004), *Neural Mechanisms for Prediction: Do Insects Have Forward Models?*, „Trends in Neurosciences" 27, 5, s. 278–82, DOI: 10.1016/j.tins.2004.03.004.

Webb Barbara (2006), *Transformation, Encoding and Representation*, „Current Biology" 16, 6, s. R184–185, DOI: 10.1016/j.cub.2006.02.034.

Webb Barbara (2008), *Using Robots to Understand Animal Behavior*, „Advances in the Study of Behavior", Vol. 38, ed. H. Jane Brockmann, Timothy J. Roper, Marc Naguib, Katherine E. Wynne-Edwards, Chris Barnard, John C. Mitani, s. 1–58, Elsevier, DOI: 10.1016/S0065-3454(08)00001-6.

Wilson R.A., Craver C.F. (2007), *Realization: Metaphysical and Scientific Perspectives*, „Philosophy of Psychology and Cognitive Science", ed. Paul Thagard, s. 81–104, North Holland, DOI: 10.1016/B978-044451540-7/50020-7.

Wimsatt William C. (2007), *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*, Cambridge, Mass.: Harvard University Press.

Woodward James (2002), *There Is No Such Thing as a Ceteris Paribus Law*, „Erkenntnis" 57, 3, s. 303–328, DOI: 10.1023/A:1021578127039.

Woodward James (2003), *Making Things Happen*, Oxford: Oxford University Press.

Streszczenie

Bronię tezy, że podstawowym rodzajem wyjaśniania w kognitywistyce jest wyjaśnianie działania mechanizmów przetwarzania informacji. Mechanizmy te stanowią złożone, zorganizowane układy, których funkcjonowanie zależy od interakcji ich części i zachodzących w nich procesów. Konstytutywne wyjaśnianie działania każdego takiego mechanizmu musi obejmować zarówno odniesienie do środowiska, w którym mechanizm występuje, jak i roli, jaką w nim odgrywa. Rolę tę tradycyjnie w kognitywistyce określa się mianem „kompetencji". Aby w pełni wyjaśnić, jak ta rola jest odgrywana, należy wyjaśnić z kolei procesy przetwarzania informacji zachodzące wewnątrz samego mechanizmu osadzonego w środowisku. Zazwyczaj wyjaśnienie na tym poziomie ma postać modelu obliczeniowego, na przykład w postaci programu komputerowego lub wytrenowanej sieci neuropodobnej. Jednak na tym poziomie wyjaśnienie się nie kończy. Do zbadania pozostaje, jak realizowany jest sam program (lub jakie procesy odpowiadają za przetwarzanie informacji w sieci neuronalnej). Na dwóch diametralnie różnych przykładach z historii kognitywistyki pokazuję, na czym polega wielopoziomowość wyjaśniania kognitywistycznego. Przykładami tymi są wyjaśnienie rozwiązywania problemów proponowane przez Simona i Newella (1972) oraz wyjaśnienie procesu fonotaksji u świerszczy proponowane przez Barbarę Webb (1995).