

Paweł Gładziejewski

Reprezentacjonizm a wyjaśnienie mechanistyczne w kognitywistyce

Filozofia Nauki 21/4, 51-77

2013

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Paweł Gładziejewski

Reprezentacjonizm a wyjaśnianie mechanistyczne w kognitywistyce

WSTĘP: REPREZENTACJONIZM I NAUKI KOGNITYWNE

Często przyjmuje się, że ludzie posługują się przednaukowym i tylko w ograniczonym stopniu uświadamianym rozumieniem umysłu, zwanym „psychologią potoczną”. Toczy się wiele sporów dotyczących jej poszczególnych aspektów, np. tego, czy (i w jakim sensie) jest wrodzona, czy przynajmniej w pewnym stopniu dysponują nią także inne naczelne, czy stanowi teorię w pewnej mierze przypominającą teorie naukowe. Powszechne jest jednak przekonanie, że psychologia potoczna pozwala nam kierować swoimi działaniami w życiu społecznym dzięki temu, że zakładamy, iż u podstaw własnych i cudzych zachowań stoją umysłowe reprezentacje świata, czyli stany mające treść intencjonalną. Paradygmatycznymi stanami tego rodzaju są przekonania i pragnienia. Tym samym psychologia potoczna stanowi pewnego rodzaju *reprezentacjonistyczną* koncepcję umysłu.

Ponadto przez dziesięciolecia w zasadzie nie kwestionowano założenia, że reprezentacjonizm nie tylko może, lecz także powinien być podstawą naukowej koncepcji umysłu, tworzonej przez przedstawicieli nauk kognitywnych. Reprezentacje umysłowe stanowić miały narzędzie eksplanacyjne zarówno dla „zwykłych” podmiotów koordynujących swoje działania w świecie społecznym, jak i dla naukowców dążących do sformułowania całkowicie naturalistycznego wyjaśnienia funkcji, własności czy procesów o charakterze umysłowym bądź poznawczym. Przyjęło się wręcz uznawać reprezentacjonizm za cechę definicyjną kognitywistyki jako przedsięwzięcia naukowego, a nie jedynie za jedną z możliwości teoretycznych w obrębie tej dziedziny (por. np. Thagard 1996). Jednocześnie sposób rozumienia reprezentacji w naukach kognitywnych mógł w zasadniczy sposób różnić się od tego, jak pojęcie

to rozumiane jest w psychologii potocznej. Struktury nazywane przez kognitywistów reprezentacjami często w zasadniczy sposób różniły się od zdroworozsądkowych przekonań i pragnień pod względem posiadanych przez siebie własności semantycznych lub funkcjonalnych, a także pod względem zakresu eksplanandów, do których wyjaśnienia miały służyć (por. Cummins 1989, Godfrey-Smith 2004). Nie zmienia to jednak faktu, że pojęciu reprezentacji przyznaje się w kognitywistyce — a w każdym razie przyznawało się do niedawna — centralną pozycję.

Zgodnie z przyjętym w artykule rozumieniem reprezentacjonizm to pogląd, w myśl którego pojęcie reprezentacji umysłowej w naukach kognitywnych ma do spełnienia fundamentalną (w mocnej wersji — absolutnie niezbywalną) rolę eksplanacyjną. Bez postulowania reprezentacji nie uda się wyjaśnić żadnego lub większości spośród zjawisk interesujących kognitywistę. Rzecz jasna tak rozumianemu „eksplanacyjnemu” reprezentacjonizmowi towarzyszy zazwyczaj ontologiczna teza, zgodnie z którą reprezentacje istnieją — są realnymi, wewnętrznymi strukturami odgrywającymi rolę przyczynową w kierowaniu zachowaniami organizmów, a przynajmniej organizmów posiadających umysł. Można też przyjąć, że w praktyce reprezentacjonizm ontologiczny przyjmuje się na podstawie (założenia o) adekwatności reprezentacjonizmu eksplanacyjnego.

Głównym przedmiotem artykułu będzie reprezentacjonizm eksplanacyjny. Podjęty zostanie temat konsekwencji, które dla tak rozumianego reprezentacjonizmu ma akceptacja twierdzenia, że wyjaśnianie w kognitywistyce stanowi formę wyjaśniania mechanistycznego, tj. wyjaśniania zjawisk przez odkrywanie ich mechanizmów. Na czym więc mogłaby polegać rola eksplanacyjna reprezentacji, jeśli założymy, że nauki kognitywne wyjaśniają interesujące je zjawiska mechanistycznie? Czym mogłyby być mechanizmy reprezentacyjne, w odróżnieniu od tych, które reprezentacyjne nie są? Jak możemy rozstrzygnąć, czy postulowany przez badacza mechanizm w sposób uprawniony odwołuje się do reprezentacji umysłowych? Przedstawione tu rozważania mają na celu udzielenie przynajmniej częściowych odpowiedzi na te pytania. Wymienione kwestie są obecnie szczególnie istotne, jako że w ciągu dwóch ostatnich dekad coraz większe wpływy na pograniczu filozofii i nauk kognitywnych zyskuje stanowisko antyreprezentacjonistyczne, towarzyszące często tzw. ucieleśnionym i enaktywnym koncepcjom poznania i umysłu. Autorzy broniący ujęcia antyreprezentacjonistycznego *również* odwołują się do eskplanacyjnego statusu reprezentacji w naukach kognitywnych, przekonują jednak, że reprezentacje w jakimkolwiek teoretycznie ciekawym sensie nie są potrzebne do wyjaśniania zjawisk kognitywnych (van Gelder 1995, Chemero 2009). Dokładne i dokonane *explicite* określenie, kiedy konkretne wyjaśnienie jest reprezentacyjne (przy założeniu, że w ogóle może takie być), jest krokiem w kierunku sformułowania kryteriów, według których powinien być rozstrzygany spór między reprezentacjonistami a antyreprezentacjonistami.

Pierwsza część artykułu poświęcona jest koncepcji wyjaśniania mechanistycznego, ze szczególnym uwzględnieniem specyfiki mechanistycznych wyjaśnień formułowanych w kognitywistyce. W części drugiej został krótko przedstawiony wpływ

przyjęcia perspektywy mechanistycznej na zagadnienie reprezentacjonizmu. Część trzecia omawia kryteria oceny, czy konkretne mechanistyczne wyjaśnienie określonego zjawiska kognitywnego jest — w prawomocny i nietrywialny sposób — reprezentacyjne. Zgodnie z przedstawioną tu propozycją kryterium takiego dostarcza sformułowana przez Williama Ramseya (2007) strategia oparta na tzw. „wymogu opisu zadań”. W części czwartej wysunięta zostanie hipoteza dotycząca wyjaśnień reprezentacyjnych: wyjaśnienia reprezentacyjne, które czynią zadość wymaganiom mechanicyzmu, to wyjaśnienia odwołujące się do reprezentacji umysłowych jako ugruntowanych w strukturalnym podobieństwie między samymi reprezentacjami (nośnikami reprezentacji) a tym, co one reprezentują. Innymi słowy, wyjaśnienia zjawisk kognitywnych odwołujące się do wewnętrznych, neuronalnych/obliczeniowych modeli lub symulacji środowiska zewnętrznego mogą być i na ogół są zarówno mechanistyczne, jak i reprezentacyjne.

1. WYJAŚNIANIE MECHANISTYCZNE W KOGNITYWISTYCE: PODSTAWOWE ZAŁOŻENIA

Robert Cummins (2000) zwrócił uwagę na fakt, że klasyczny, dedukcyjno-nomologiczny model wyjaśniania naukowego jest w zasadniczy sposób nieadekwatny, jeśli chcemy go potraktować jako opis zasad wyjaśniania zjawisk w naukach kognitywnych. Jest tak dlatego, że prawa odkrywane przez badaczy zajmujących się umysłem nie spełniają w istocie żadnej roli eksplanacyjnej, podczas gdy model dedukcyjno-nomologiczny związany jest z ideą wyjaśniania jako subsumpcji pod prawa. Podciągnięcie pod określone prawo psychologiczne nie jest traktowane przez kognitywistów jako zadowalające wyjaśnienie żadnego zjawiska umysłowego. Zamiast tego, jak przekonuje Cummins, prawa uznawane są raczej za *opis* fenomenów, które mają dopiero podlegać wyjaśnieniu¹. Czym jednak może być wyjaśnianie w kognitywistyce, jeśli nie polega na podciąganiu pod prawa?

Coraz bardziej wpływowa, a jednocześnie wyraźnie alternatywna wobec modelu dedukcyjno-nomologicznego, jest teoria oparta na idei wyjaśniania mechanistycznego. Jej zasadnicze założenia sformułowane zostały jeszcze w latach siedemdziesiątych poprzedniego wieku w pracach Cumminsa (1975) i Williama Wimsatta (1974).

¹ Cummins (2000) podaje tu przykład efektu McGurka, czyli zjawiska dotyczącego interakcji różnych modalności słyszalnych. Efekt ten polega na tym, że percepcja słuchowa sylab przez daną osobę podlega systematycznym zmianom, jeśli osoba słuchająca widzi jednocześnie osobę wypowiadającą sylabę inną niż sylaba słyszana. Jak przekonuje Cummins, efekt McGurka — tak samo jak ogromna ilość innych „efektów” odkrywanych przez psychologów i kognitywistów — może zostać scharakteryzowany jako określonego rodzaju prawo psychologiczne. Powszechnie przyjmuje się jednak, że prawo to ma status eksplanandum, a nie eksplanansu. Efekt McGurka w praktyce poznawczej naukowców traktowany jest nie jako podstawa do wyjaśniania jakiegokolwiek zjawiska, lecz raczej opis zjawiska, które domaga się wyjaśnienia w kategoriach przetwarzania modalnie specyficznych informacji słuchowych i wzrokowych.

Od tamtej pory została ona w wielu aspektach uzupełniona i zyskała powszechniejszą akceptację jako model adekwatnie lub w przybliżeniu adekwatnie ujmujący naturę praktyki eksplanacyjnej naukowców pracujących w tych dziedzinach — w tym w kognitywistyce — względem których nieadekwatny deskryptywnie wydaje się model nomologiczno-dedukcyjny (por. m.in. Bechtel 2008, Bechtel, Abrahamsen 2005, Bechtel, Richardson 1993, Craver 2007, Cummins 2000, Glennan 2002, Machamer, Darden, Craver 2011)². Istnieje cały szereg subtelnych różnic w tym, jak koncepcja wyjaśniania mechanistycznego formułowana jest przez poszczególnych autorów. Przedstawiona tu rekonstrukcja ma być możliwie ogólna, tak aby uchwycić wspólne cechy tych sformułowań.

W koncepcji wyjaśniania mechanistycznego podstawowe znaczenie przyznawane jest pojęciu mechanizmu. Oto dwie przykładowe, bardzo ogólne, a jednocześnie zbliżone do siebie treściowo definicje mechanizmu:

Mechanizm to struktura pełniąca pewną funkcję ze względu na swoje komponenty, operacje komponentów oraz ich organizację. Skoordynowane funkcjonowanie mechanizmu jest odpowiedzialne za powstanie jednego lub wielu różnych zjawisk (Bechtel, Abrahamsen 2005: 423).

Mechanizmy to jednostki [*entities*] oraz działania [*activities*] zorganizowane tak, że razem egemplifikują one zjawisko stanowiące eksplanandum (Craver 2007: 6).

Mechanizm jest zatem złożonym układem, którego wewnętrzna budowa oraz organizacja jest odpowiedzialna za istnienie czy zachodzenie wyjaśnianego zjawiska. Mechanizmy w tym znaczeniu mogą mieć zarówno charakter naturalny (biologiczny, astrofizyczny, chemiczny itd.), społeczny (np. grupy, instytucje), jak i być artefaktami (dowolny układ mechaniczny pełniący określone funkcje). Mówienie o mechanizmie jako dobrze wyodrębnionym złożonym *obiekcie* może być jednak mylące. Mechanizmy w sensie, o którym tu mowa, indywiduowane są częściowo ze względu na *zjawiska*, które mają wyjaśniać (por. Craver 2007). Dlatego też powinniśmy myśleć o mechanizmach w kategoriach „mechanizmów czegoś”. Odwołując się do przykładów biologicznych, możemy mówić o mechanizmach syntezy białek, powstawania komórek, wydzielania neuroprzekaźników, reprodukcji lub rozwoju całych gatunków. W jednym złożonym obiekcie (np. organizmie żywym lub pojedynczej komórce) możemy zatem wyróżnić wiele mechanizmów odpowiadających za odmienne zjawiska. Przez „zjawiska” stanowiące eksplananda wyjaśniania mechanistycznego rozumie się funkcje, procesy, własności i zdolności pewnego złożonego obiektu bądź układu. To oczywiście bardzo ogólna i ontologicznie heterogeniczna charakterystyka eksplanandów wyjaśniania mechanistycznego. Niektórzy autorzy próbują ją ujednoclić i doprecyzować, charakteryzując eksplananda jako regularne korelacje między stanami początkowymi mechanizmu (na wejściu) a stanami końcowymi (na wyjściu) (por. Glennan 2002, Machamer, Darden, Craver 2011).

² Należy wyraźnie zaznaczyć, że teoria, o której mowa, zasadniczo różni się od koncepcji przyczynowo-mechanistycznej Salmona. Różnice pomiędzy nimi omawia szerzej Glennan (2002).

Złożoność mechanizmów ma trzy aspekty. Po pierwsze, na mechanizm składają się komponenty czy też jednostki składowe³. Po drugie, na mechanizm składają się operacje lub działania wykonywane przez komponenty. Operacjami często przypisywanymi komponentom mechanizmów naturalnych są chociażby: „pompowanie”, „blokowanie”, „przyciąganie”, „otwieranie”, „filtrowanie”, „wydzielanie”. Choć niektórzy autorzy odróżniają operacje komponentów od ich *funkcji* (por. np. Machamer, Darden, Craver 2011), to na potrzeby tego artykułu przyjmuje się rozumienie operacji właśnie jako funkcjonalnego wkładu danego komponentu w działanie mechanizmu jako całości (por. Craver 2008)⁴. Po trzecie, komponenty i ich operacje wykazują określoną organizację w wymiarze czasowym, przestrzenno-strukturalnym oraz przyczynowym. Organizacja wielu naturalnych, nieagregatywnych mechanizmów jest złożona, nieliniowa i oparta na sprzężeniach zwrotnych. W przypadku tego rodzaju mechanizmów lokalizowanie komponentów oraz adekwatne przypisywanie im operacji stanowi dla badaczy zasadnicze wyzwanie poznawcze (por. Bechtel, Richardson 1993). Z tego też powodu niektórzy autorzy starają się uzgodnić mechanicyzm z teorią systemów dynamicznych, pokazując, że modele dynamiczne określonych zjawisk (np. biologicznych) mogą być pod pewnymi warunkami traktowane jako formy opisu mechanizmów odpowiedzialnych za te zjawiska (por. Bechtel, Abrahamsen 2010).

Zgodnie z koncepcją wyjaśniania mechanistycznego wyjaśnianie zjawiska stanowiącego eksplanandum polega na dekompozycji badanego systemu celem odkrycia i opisanego mechanizmu odpowiedzialnego za zachodzenie tego zjawiska przy jednoczesnym pokazaniu, *w jaki sposób* działanie mechanizmu umożliwia jego zachodzenie. Wyjaśnienie zjawiska polega więc na pokazaniu, jak jest ono umożliwiane przez zorganizowane działanie komponentów wykonujących określone operacje. Odwołując się do prostszego przykładu biologicznego, wyjaśnienie pompowania krwi przez ludzkie serce polegałoby na pokazaniu, jak pompowanie jest umożliwiane przez cechujący się określoną organizacją czasową i przestrzenną układ skurczają-

³ Craver (2007) szczegółowo omawia problem, jak należy w kontekście wyjaśniania mechanistycznego rozumieć relację zachodzącą między mechanizmem a jego komponentami. Odrzuca między innymi możliwość, zgodnie z którą byłaby to relacja agregacji (własności całości stanowią sumę, czyli agregat własności jej części) bądź relacja zawierania przestrzennego i proponuje uznać, że poziomy organizacji mechanizmów powiązane są nieekstensjonalną i przeciwzrotną relacją kompozycji: na wyższym poziomie mamy do czynienia z mechanizmem jako całością (i cechującą go jako całość zdolnością), a na poziomie niższym — z częściami owego mechanizmu, których zorganizowane operacje (działania) umożliwiają zachodzenie zjawiska (zdolności) wyróżnianego na poziomie wyższym.

⁴ Mówiąc o funkcjach w tym znaczeniu, nie ma się na myśli funkcji właściwych w sensie Millikan, tj. funkcji zawdzięczających swój status własnej historii ewolucyjnej. Charakteryzowanie komponentów mechanizmów jako pełniących role funkcjonalne bądź jako funkcjonalnie zależnych od innych komponentów nie niesie tu zatem zobowiązań dotyczących filogenezy gatunków lub ich cech. Przypisywanie danemu komponentowi mechanizmu funkcji wymaga jedynie usytuowania go w ramach szerszej organizacji mechanizmu jako całości (por. Craver 2008).

cych i rozkurczających się (operacje) komór (komponenty). Przykład ten pozwala też zilustrować *hierarchiczną* naturę mechanizmów. Z jednej strony na wyższym poziomie, serce jako takie stanowi komponent mechanizmu obiegu krwi. Z drugiej strony operacje komponentów serca (skurcze i rozkurcze komór) same mogą zostać wyjaśnione przez pokazanie, jak umożliwiające są one przez mechanizmy z niższego poziomu⁵.

Mechanicystyczne podejście do wyjaśniania naukowego ma istotne konsekwencje i ciekawe własności, w tym takie, które wyraźnie odróżniają tę koncepcję od modelu nomologiczno-dedukcyjnego. Jedną z różnic polega na tym, że w porównaniu do modelu dedukcyjno-nomologicznego w koncepcji mechanistycznej rola teorii oraz praw jest marginalna. Kolejna (choć ściśle związana z pierwszą) różnica związana jest z tym, że w perspektywie mechanicyzmu eksplanans nie musi mieć i na ogół nie ma postaci rozumowania dedukcyjnego. Rolę wyjaśnień pełni to, co pozwala badaczowi zrozumieć, jak badane zjawisko powstaje w wyniku działania pewnego mechanizmu. Dlatego też za pełnoprawne eksplanansy można uznać adekwatnie oddające działanie odpowiednich mechanizmów diagramy bądź modele (por. Bechtel 2008, Bechtel, Abrahamsen 2005).

Należy jednak zwrócić uwagę, że choć wyjaśnianie zjawisk przez odwołanie do mechanizmów stanowi praktykę epistemiczną, to warunkiem poprawności wyjaśnienia jest, by opisywany w ten czy inny sposób mechanizm był rzeczywiście odpowiedzialny za zachodzenie tego zjawiska w świecie fizycznym (por. Bechtel 2008, Craver 2007). Na przykład model matematyczny zjawiska, który dobrze ujmuje korelacje wejście—wyjście, nie jest wyjaśnieniem, dopóki nie pokażemy, że odzwierciedla także procesy zachodzące *wewnątrz* mechanizmu odpowiedzialnego za powstanie danego zjawiska. Dlatego też mechanistyczny model wyjaśniania nie poddaje się jednoznacznej interpretacji w świetle wprowadzonego przez Salmona rozróżnienia między wyjaśnieniami ontycznymi i epistemicznymi (Bechtel, Abrahamsen 2005).

Koncepcja wyjaśniania mechanistycznego pozwala także nadać sens intuicji, zgodnie z którą sukces eksplanacyjny jest czymś stopniowalnym. Należy bowiem odróżnić kompletne opisy, diagramy czy modele mechanizmów od takich, które co prawda pozwalają częściowo wyjaśnić eksplanandum, ale są zaledwie niepełnymi szkicami (por. Craver 2007). Warto wreszcie dodać, że mechanicyzm pozwala przeformułować wiele klasycznych problemów filozofii nauki oraz metafizyki, np. problem redukcji i emergencji (w tym problem możliwości uzgodnienia redukcjonizmu z emergentyzmem), natury przyczynowości (w tym przyczynowości odgórnjej), relacji między przewidywaniem a wyjaśnianiem, zmiany naukowej i jedności nauki (por. np. Craver, Bechtel 2006, Glennan 2002, Machamer, Darden, Craver 2011).

⁵ Mówienie o poziomach w tym kontekście nie dotyczy ontycznych poziomów natury jako takiej, lecz jest zrelatywizowane do badanego mechanizmu. X wykonujące operację (działanie) ϕ jest na niższym poziomie niż Y wykonujące operację ψ w tw. X wykonujące operację ϕ stanowi komponent mechanizmu odpowiadającego za wykonywanie operacji ψ przez Y (Craver 2007).

Coraz powszechniej uważa się, że do grupy dziedzin, w których praktyka eksplanacyjna polega na odkrywaniu i opisywaniu mechanizmów, należy także kognitywistyka. Wyjaśnienia kognitywistyczne nie opierają się na dedukowaniu zjawisk na podstawie praw, lecz zmierzają do udzielenia odpowiedzi na pytanie, „jak działa” system poznawczy, tj. jakie mechanizmy umożliwiają mu egzemplifikowanie określonych własności (por. Cummins 2000, Bechtel 2008). Podlegającymi wyjaśnianiu zjawiskami są fenomeny o charakterze kognitywnym i umysłowym. Cummins (2000) zwraca uwagę na fakt, że głównymi eksplanandami nauk kognitywnych są *zdolności*, które przypisujemy systemowi poznawczemu jako całości zaangażowanej w określone rodzaje interakcji ze swoim środowiskiem (naturalnym, społecznym, a w wypadku umysłów ludzkich także kulturowym). Zdolnościami w tym sensie są chociażby: *mind-reading* („czytanie w myślach”, czyli przypisywanie stanów umysłowych sobie i innym), percepcja kategoriałna, multimodalna integracja sensoryczna, formowanie pojęć i podejmowanie decyzji. Cummins proponuje traktować zdolności jako własności dyspozycyjne, które można charakteryzować za pomocą okreśów warunkowych określających zachowanie systemu poznawczego przy zająciu określonych warunków⁶.

Wyjaśnianie w kognitywistyce polega zaś na odkrywaniu mechanizmów zdolności. Mechanizmy te są realnymi strukturami (tzn. zorganizowanymi komponentami wykonującymi określone operacje) zawartymi w biologicznym mózgu. Nawet jeśli operacje przypisywane komponentom tych mechanizmów polegają na przetwarzaniu informacji, a w związku z tym ważną rolę pełnią w kognitywistyce *obliczeniowe* modele zdolności poznawczych, to o poprawnym wyjaśnianiu może być mowa tylko wtedy, gdy modele te są *neurobiologicznie adekwatne*: kiedy reprezentują sposób, w jaki informacje przetwarzane są w rzeczywistych mózgach⁷. Neurobiologicznie nierealistycznym modelom obliczeniowym zdolności poznawczych nie może więc przysługiwać status wyjaśnień. Analogicznie jest w wypadku diagramów funkcjonalnej organizacji systemu poznawczego, bardzo popularnych do niedawna w psychologii poznawczej. Z perspektywy mechanicyzmu diagramy takie mają wartość, o ile odzwierciedlają strukturę realnych mechanizmów stojących u podstaw zdolności poznawczych. Najczęściej są na tyle nieprecyzyjne, że przysługuje im raczej status szkiców mechanizmów niż kompletnych wyjaśnień mechanistycznych (por. Piccinini, Craver 2011).

⁶ Odkrywanie i charakteryzowanie zdolności samo w sobie stanowi wyzwanie naukowe. Często dopiero po przeprowadzeniu wielu badań empirycznych okazuje się, że zdolność traktowana przednaukowo jako jedna (np. pamięć) w istocie rozpada się na kilka w dużym stopniu niezależnych zdolności (pamięć krótkotrwała, długotrwała, epizodyczna, ikoniczna, przestrzenna itd.), a za każdą z nich odpowiada inny mechanizm.

⁷ Należy jednak pamiętać, że nadanie sensu idei, zgodnie z którą neuronalne mechanizmy stojące u podstaw zdolności poznawczych w ogóle mają — w jakimś nietrywialnym sensie — charakter obliczeniowy samo w sobie jest znaczącym problemem filozoficznym i naukowym (por. np. Miłkowski 2013, Piccinini 2007).

Zarówno w kognitywistyce, jak i w innych dziedzinach wyjaśnienia mechanistyczne są hierarchiczne. System poznawczy i przysługujące mu zdolności znajdują się na wyższym poziomie organizacji niż zorganizowane komponenty i operacje składające się na mechanizmy stojące u podstaw systemu. Rozważmy mechanistyczne wyjaśnienie pamięci przestrzennej, wskazywane przez Bechtela (2008) i Cravera (2007) jako jeden z wzorcowych przypadków zastosowania strategii mechanistycznej do wyjaśnienia zjawiska kognitywnego. Na poziomie systemu poznawczego mamy w tym wypadku pamięć przestrzenną, rozumianą jako zdolność do pamiętania swojego otoczenia przestrzennego oraz sterowania dzięki temu swoimi działaniami w tym otoczeniu. Dostarczenie dokładnego opisu tej kompetencji jest samo w sobie wyzwaniem naukowym, które wymaga badań eksperymentalnych (najczęściej stosowane są tu badania zachowań szczurów w stworzonych na potrzeby eksperymentu labiryntach, por. Craver 2007). Wyjaśnienie pamięci przestrzennej wymaga jednak zejścia na niższy poziom przez dostarczenie opisu mechanizmu, który za nią odpowiada. W tym wypadku mechanizm ten opiera się na tworzeniu wewnętrznych map przestrzennych środowiska w hipokampie, a także skroniowych i czołowych strukturach korowych. Chociaż mapy stanowią tu eksplanans, to jednocześnie są także eksplanandum, które wyjaśniane jest za pomocą mechanizmu z jeszcze niższego poziomu. Jest to mianowicie mechanizm długotrwałego wzmocnienia synaptycznego neuronów hipokampu, który to sam wyjaśniony zostać może za pomocą mechanizmu z jeszcze niższego (molekularnego) poziomu organizacji układu nerwowego.

Co ciekawe, „hierarchiczna” koncepcja wyjaśniania mechanistycznego pozwala zinterpretować wprowadzone przez Dennetta (1995), klasyczne już odróżnienia na osobowy i subosobowy poziom opisywania i wyjaśniania zjawisk umysłowych. Co prawda rozróżnieniem tym posługuje się wielu współczesnych autorów, lecz często w nieprecyzyjny i za mało techniczny sposób. Z perspektywy mechanicyzmu intencjonalny, osobowy poziom odwoływałby się do systemów poznawczych jako całości zaangażowanych w wielorakie interakcje ze swoim środowiskiem, natomiast poziom subosobowy (precyzyjniej: *poziomy* subosobowe) byłby poziomem opisu wewnętrznej struktury mechanizmów, które umożliwiają zjawiska z poziomu osobowego (por. Gładziejewski 2012, gdzie szerzej omawia się tę ideę i niektóre jej konsekwencje).

Dysponując ogólnym rozumieniem natury wyjaśniania mechanistycznego, możemy powrócić do głównego zagadnienia. Czy postawienie problemu reprezentacji w kontekście teorii wyjaśniania mechanistycznego pozwala nam przybliżyć się do rozwiązania tego problemu? Czy w perspektywie mechanicyzmu reprezentacje umysłowe nadal mogą stanowić dla kognitywistów wartościowe narzędzie eksplanacyjne? A jeśli tak, to jak należy rozumieć „reprezentacje” i „reprezentowanie”?

2. WYJAŚNIANIE MECHANISTYCZNE A REPREZENTACJONIZM

Koncepcja wyjaśnienia mechanistycznego powinna jednoznacznie przekładać się na zagadnienie reprezentacjonizmu, tj. powinna umożliwić sformułowanie odpowiedzi na pytanie o to, kiedy reprezentacje umysłowe coś wyjaśniają i na czym to wyjaśnianie polega. Innymi słowy, mechanicyzm powinien, jako teoria wyjaśniania, umożliwić sformułowanie kryteriów, według których możemy w kognitywistyce przyznać reprezentacjom wartość eksplanacyjną. Z mechanistycznego punktu widzenia wyjaśnianie w naukach kognitywnych polega na odkrywaniu mechanizmów. Jego celem jest pokazanie, jak odpowiednio zorganizowane komponenty mechanizmu i realizowane przez nie operacje umożliwiają zajście wyjaśnianego zjawiska kognitywnego. Jeśli zatem coś ma pełnić rolę eksplanacyjną względem tego zjawiska, to powinniśmy oczekiwać, że będzie ono stanowić element mechanizmu wyjaśniającego to zjawisko. To z kolei znaczy, że będzie komponentem tego mechanizmu bądź operacją wykonywaną przez któryś z jego komponentów. Jeśli więc chcemy, aby pojęcie reprezentacji mogło odgrywać rolę w mechanistycznych wyjaśnieniach formułowanych przez kognitywistów, to powinniśmy oczekiwać, że istnieją mechanizmy, których komponentami są reprezentacje. Komponenty te powinny wykonywać w obrębie mechanizmu operację reprezentowania. Tym samym, jeśli wyjaśnienie ma być reprezentacyjne, to komponent lub komponenty postulowanego w nim mechanizmu powinny *służyć za* bądź *funkcjonować jako reprezentacje* w ramach tego mechanizmu.

Oczywiście konstatacje te mają jedynie wstępny charakter. Wciąż nie jest jasne, na czym dokładnie miałyby polegać realizowanie przez komponent mechanizmu operacji reprezentowania. Problemowi temu poświęcone są dalsze części artykułu. Założmy jednak, że mechanizmy reprezentacyjne — a w związku z tym reprezentacyjne wyjaśnienia zjawisk kognitywnych — są możliwe. Takie spojrzenie na problem reprezentacjonizmu pozwala zmodyfikować lub postawić w nowym świetle przynajmniej niektóre szczegółowe zagadnienia ściśle powiązane z tym problemem. Niżej przedstawione są cztery bardzo ogólne konsekwencje przyjęcia perspektywy mechanistycznej w kontekście sporu reprezentacjonizm—antyreprezentacjonizm. Lista ta zapewne nie jest kompletna, można jednak przyjąć, że zawiera konsekwencje najbardziej istotne.

(a) Funkcjonowanie jako reprezentacja a zagadnienie treści intencjonalnej

Jedną z zasadniczych konsekwencji przyjęcia mechanistycznej perspektywy względem problemu reprezentacji jest fakt, że w ten sposób dość naturalnie przenosimy uwagę z zagadnienia treści intencjonalnej na zagadnienie funkcjonowania jako reprezentacja w obrębie systemu poznawczego. W ciągu ostatnich dziesięcioleci naturalistycznie nastawionych filozofów umysłu zaprzętało pytanie o to, jakiego rodzaju *stricte* naturalne własności (w tym własności relacyjne) stanów neuronalnych sprą-

wiają (w znaczeniu konstytutywnym, a nie przyczynowym), że stany te posiadają treści intencjonalne, które normalnie przypisuje się stanom postulowanym w psychologii potocznej. Udzieleniu odpowiedzi na to pytanie służyły często złożone eksperymenty myślowe, których znaczenie dla realnej praktyki badawczej kognitywistów było raczej marginalne. Jednym z głównych punktów spornych tych dyskusji była opozycja między teoriami internalistycznymi a eksternalistycznymi. Wyniki rozważań tego rodzaju na ogół nie przekładają się jednak jednoznacznie na kwestię tego, co to znaczy pełnić rolę reprezentacji w ramach mechanizmu kognitywnego, tj. nie przekładają się na zagadnienie funkcjonalnych kryteriów bycia reprezentacją. Na przykład na posiadanie tego rodzaju reprezentacyjnej roli funkcjonalnej nie przekłada się w żaden sposób fakt, że dany egzemplarz bądź typ stanu neuronalnego lub obliczeniowego znajduje się w przyczynowo-nomologicznej relacji do występowania określonych okoliczności środowiskowych⁸. Tymczasem w myśl jednej z wpływowych koncepcji filozoficznych to właśnie istnienie tej przyczynowo-nomologicznej relacji miałyby nadawać treść reprezentacjom. Jeśli więc interesuje nas nie tyle naturalizacja treści, ile zagadnienie, kiedy reprezentacje mają dla kognitywistów realną wartość eksplanacyjną, to powinniśmy przenieść ciężar rozważań z problemu treści na w znacznej mierze odrębny i niezależny problem funkcjonowania jako reprezentacja⁹.

⁸ To, że jakiś stan x jest nomicznie współzmienny z zajściem okoliczności środowiskowej y , nie nam jeszcze nie mówi o tym, jak x funkcjonuje (jakie spełnia role przyczynowe) w ramach mechanizmu, którego jest komponentem; por. O'Brien, Opie 2004.

⁹ Należy poczynić tu trzy istotne zastrzeżenia. Po pierwsze, niektóre koncepcje powstałe w kontekście problemu treści mogą być równie dobrze potraktowane jako teorie dające pewnego rodzaju odpowiedź na pytanie o to, co to znaczy funkcjonować jako reprezentacja. Np. Ramsey (2007) przekonuje, że tak właśnie można traktować teorie, które do standardowego ujęcia przyczynowo-nomologicznego dodają elementy o charakterze teleologicznym. Po drugie, odróżnienie problemu naturalizacji treści od problemu funkcjonalnych kryteriów bycia reprezentacją nie implikuje w żadnym wypadku, że powinniśmy przyjąć „wąską” koncepcję treści, zgodnie z którą treść stanów wewnętrznych jest konstytuowana przez ich role funkcjonalne. Chodzi raczej o to, że obydwa zagadnienia można traktować oddzielnie. Propozycje dotyczące tego, jakie własności (relacyjne lub nie) nadają treść intencjonalną, na ogół nie będą przesądzały o tym, w jaki sposób stany posiadające tę treść funkcjonują w ramach określonego mechanizmu, ani nawet o tym, czy ich role funkcjonalne są w jakimś nietrywialnym, wartościowym eksplanacyjnie sensie reprezentacyjne. Po trzecie, oba problemy nie muszą być — i prawdopodobnie nie będą — całkowicie odrębne. Mówiąc, że pewien stan wewnętrzny pełni rolę funkcjonalną jako reprezentacja, będziemy mieć między innymi na myśli to, że rola ta jest (współ)determinowana przez treść tego stanu. Funkcjonować jako reprezentacja to przecież pełnić takie, a nie inne role dlatego, że posiada się taką, a nie inną treść (por. Ramsey 2007). W takim wypadku problem funkcjonowania jako reprezentacja będzie odrębny, ale nie całkowicie niezależny od pytania o to, co nadaje stanom wewnętrznym treść.

(b) Spór o reprezentacjonizm a założenia dotyczące natury wyjaśniania w kognitywistyce

Jak zostało zaznaczone na wstępie, stanowisko reprezentacjonistyczne oparte jest na idei, że pojęcie reprezentacji ma zasadnicze znaczenie dla praktyki eksplanacyjnej kognitywistów, a nawet jest ono dla niej niezbędne. Analogicznie antyreprezentacjonizm może być rozumiany jako stanowisko, zgodnie z którym pojęcie reprezentacji jest dla kognitywistyki eksplanacyjnie zbędne. Pojęcie wyjaśniania jest więc podstawowe w sporze między reprezentacjonistami a antyreprezentacjonistami. Niestety wydaje się, że przedstawiciele obydwu stron zbyt rzadko czynią starania, aby przed rozstrzygnięciem kwestii reprezentacji umysłowych odpowiedzieć na pytania o to, (1) czym dokładnie jest wyjaśnianie w kognitywistyce; (2) czym są lub mogłyby być wyjaśnienia reprezentacyjne w odróżnieniu od niereprezentacyjnych; (3) czy wyjaśnienia reprezentacyjne (w jakimś wartościowym poznawczo sensie terminu „reprezentacyjny”) są w ogóle możliwe. Spojrzenie na problem reprezentacjonizmu z perspektywy mechanicyzmu powinno przyczynić się do rozwiązania wszystkich tych problemów, a tym samym pozwolić na dookreślenie kryteriów, wedle których spór między stronami powinien być rozstrzygany. Ponadto przyjęcie określonych założeń dotyczących wyjaśniania w kognitywistyce może pomóc wykryć błędy i nieuprawnione założenia tkwiące w niektórych argumentacjach. Np. antyreprezentacjonści powołują się często na, według nich, z zasady niekompatybilne z reprezentacjonizmem rozumienie umysłu jako sprzężonego ze środowiskiem systemu opisywanego w języku teorii systemów dynamicznych (van Gelder 1995, Chemero 2009). Tymczasem okazuje się, że modele dynamiczne można traktować jako opisy mechanizmów; co więcej, niewykluczone, że dopiero potraktowane jako opisy mechanizmów modele te mogą cokolwiek wyjaśniać (por. Kaplan, Bechtel 2011). Jeśli więc możemy sformułować koncepcję pokazującą, pod jakimi warunkami wyjaśnienie mechanistyczne jest reprezentacyjne, a wyjaśnienia odwołujące się do modeli dynamicznych są (przy określonych uzupełnieniach) wyjaśnieniami mechanistycznymi, to można sądzić, że wyjaśnienia odwołujące się do modeli dynamicznych *mogą* być reprezentacyjne. Zatem bycie zwolennikiem dynamistycznego ujęcia umysłu i poznania nie przesądza jeszcze o byciu antyreprezentacjonistą.

(c) Reprezentacjonizm globalny i lokalny

Zostało już wspomniane, że mechanizmy mogą być po części indywiduowane przez odniesienie do zjawisk, które mają wyjaśniać. Być może często bardziej adekwatne niż mówienie o mechanizmie *per se* jest mówienie o mechanizmie *czegoś* (mitozy, syntezy białek, powstawania supernowych, kategoryzacji percepcyjnej itd.). Różne zjawiska wyjaśniane są na ogół za pomocą odmiennych mechanizmów. Oczywiście idea ta stosuje się również do kognitywistyki. Konkretnie eksplananda nauk

kognitywnych będą często wyjaśniane za pomocą odrębnych mechanizmów. Prawdopodobnie tylko część spośród tych eksplanandów może zostać wyjaśniona przez odwołanie do mechanizmów, których komponenty *reprezentują* środowisko zewnętrzne. Które zjawiska domagają się wyjaśnienia w terminach reprezentacyjnych, a które nie, stanowi odrębny problem, w dużej mierze nierozwiązywalny na gruncie rozważań czysto pojęciowych. Być może najważniejszym do tej pory głosem w tej sprawie był opublikowany jeszcze w latach dziewięćdziesiątych poprzedniego wieku artykuł Clarka i Toribia (1994), w którym autorzy zaproponowali szereg kryteriów służących do odróżniania fenomenów kognitywnych, które są reprezentacyjnie wymagające (*representation-hungry*) od tych, które takie nie są. Abstrahując jednak od tego szczególnego zagadnienia, należy podkreślić kwestię ogólniejszą: w kontekście mechanicyzmu spór reprezentacjonizm—antyreprezentacjonizm powinien być rozstrzygany w odniesieniu do *poszczególnych eksplanandów* (lub kategorii eksplanandów). Niewykluczone, że poprawne wyjaśnienie niektórych zjawisk kognitywnych będzie odwoływało się do pojęcia reprezentacji, a innych — nie. W każdym razie przyjęcie postawy mechanistycznej pozwala sądzić, że spór reprezentacjonistów z antyreprezentacjonistami nie powinien być rozumiany jako spór globalny, dotyczący kognitywistyki jako całości, lecz lokalny, tj. jako spór o wyjaśnienia konkretnych zjawisk kognitywnych. W takim wypadku jest się nie tyle (anty)reprezentacjonistą po prostu, ile raczej (anty)reprezentacjonistą w stosunku do konkretnego zjawiska.

(d) Reprezentacja jako eksplanans i eksplanandum

Ostatnia konsekwencja przyjęcia perspektywy mechanistycznej jest być może filozoficznie najważniejsza. Zostanie jej też poświęcone nieco więcej uwagi niż pozostałym. Mimo to wyczerpujące jej omówienie nie jest tutaj możliwe ze względów objętościowych, a poczynione uwagi należy potraktować jako wstępny i ogólnikowy szkic.

Spojrzenie na zagadnienie reprezentacjonizmu z punktu widzenia koncepcji wyjaśniania mechanistycznego pomaga zauważyć, że na pograniczu filozofii umysłu i kognitywistyki istnieją w istocie dwa „problemy reprezentacji” i związane z nimi projekty filozoficzne. Projekty te nie są na ogół odróżniane *explicite* (za ważne wyjątki można uznać Cummins 1989 i Godfrey-Smith 2004), co ma bardzo istotne i, jak się wydaje, często niekorzystne filozoficzne konsekwencje. Różnicę między tymi projektami można najprościej ująć, stwierdzając, że w jednym traktuje się reprezentacje umysłowe jako eksplanandum, a w drugim — jako eksplanans. Funkcję eksplanandum reprezentacje pełnią w projekcie naturalizacji intencjonalności, tj. w projekcie mającym na celu określenie (w kategoriach czysto naturalistycznych, nieodwołujących się do własności intencjonalnych) warunków wystarczających, by można było uznać dany typ lub egzemplarz wewnętrznego stanu neuronalnego/obliczeniowego za identyczny z danym typem lub egzemplarzem stanu intencjonalnego (por. Fodor 1987, Millikan 1984, Dretske 1988). Innymi słowy, reprezentacje umysłowe

muszą tu zostać naturalistycznie wyjaśnione, w bardzo ogólnym i niezobowiązującym znaczeniu terminu „wyjaśnienie”. Jednocześnie reprezentacje, o których mowa, to pochodzące z psychologii potocznej, przypisywane *osobom* stany intencjonalne, takie jak przekonania i pragnienia. Drugi projekt dotyczący reprezentacji umysłowych związany jest z problemem, któremu poświęcony jest ten artykuł, czyli z wartością eksplanacyjną pojęcia reprezentacji *dla kognitywistyki*. W tym wypadku reprezentacje umysłowe traktowane są przede wszystkim jako eksplanans, jako narzędzie, za pomocą którego badacze mają wyjaśniać interesujące ich zjawiska. Za reprezentacje w tym sensie często uznaje się stany czy struktury zasadniczo różniące się od stanów przypisywanych osobom, np. komórki receptorowe, wewnętrzne emulatory systemu szkieletowo-mięśniowego i środowiska albo atraktory w przestrzeniach fazowych służących do opisu zachowania systemu poznawczego.

Jak zostało już zaznaczone, oba wymienione projekty związane z zagadnieniem reprezentacji nie są na ogół wystarczająco wyraźnie odróżniane. Co więcej, wydaje się, że proponowane rozwiązania problemu reprezentacji jako eksplanandum traktowane są często tak, jak gdyby stanowiły rozwiązanie problemu reprezentacji jako eksplanansu. Filozofowie zaangażowani w rozwiązywanie tego pierwszego problemu zakładają na ogół, że naturalizacja intencjonalności polega na pokazaniu, że *wewnętrzne* stany neuronalne lub obliczeniowe (np. ciągi symboli w Fodorowskim języku myśli) mogą posiadać własności funkcjonalne i intencjonalne, dzięki którym można uznać je za tożsame ze stanami, które ludzie wzajemnie sobie przypisują, wyjaśniając i przewidując swoje działania za pomocą psychologii potocznej. Stany te miałyby *jednocześnie* odgrywać rolę w *naukowych* (kognitywistycznych) wyjaśnieniach zachowania organizmów, w tym organizmów ludzkich. Uważa się czasem, że tylko w ten sposób psychologia potoczna mogłaby zostać uprawomocniona w świetle filozoficznego naturalizmu, a przekonania i pragnienia uznane za realne. Te same ogólne założenia stoją także u podstaw niektórych klasycznych argumentów za *eliminatywizmem* w kwestii „potocznych” stanów intencjonalnych. Zgodnie z tymi argumentami, jeśli odkrywane przez kognitywistów mechanizmy nie odwołują się do stanów bądź struktur przypominających funkcjonalnie lub intencjonalnie stany postulowane w ramach psychologii potocznej, to te drugie możemy całkowicie wyeliminować z naturalistycznej teorii umysłu (por. Ramsey, Stich, Garon 1990).

Perspektywa mechanistyczna pozwala pokazać, że jest wysoce prawdopodobne, iż pomieszanie obydwu problemów reprezentacji (oraz związanych z nimi projektów filozoficznych) jest błędem. Kluczem jest tu obserwacja, że własności i zdolności, które przypisujemy systemom naturalnym jako całościom, nie są na ogół egzemplifikowane przez komponenty mechanizmów, za pomocą których wyjaśniamy działanie tych systemów. Nie należy oczekiwać zachodzenia odpowiedniości lub izomorfii między poziomami organizacji tych systemów (por. Bechtel 1994, Bechtel, Richardson 1993). Przekładając to na zagadnienia związane z umysłem, z perspektywy mechanicyzmu najlepiej potraktować psychologię potoczną jako koncepcję sformułowaną na poziomie systemów poznawczych (alternatywnie: osób lub organizmów)

pojętych jako *całości* zaangażowane w wielorakie interakcje ze swoim środowiskiem, w tym ze środowiskiem kulturowym i społecznym (por. Bechtel, Abrahamsen 1993). Z kolei reprezentacje użyteczne z punktu widzenia kognitywistów postulowane są na niższym poziomie organizacji systemów poznawczych, tj. na poziomie mechanizmów, które umożliwiają zjawiska z poziomu systemowego. Nawet jeśli komponenty mechanizmów reprezentują, to raczej *nie* w ten sposób i nie w takim znaczeniu co „potoczne” stany intencjonalne. Potoczne i naukowe pojęcia reprezentacji mają do odegrania bardzo różne role eksplanacyjne i nie powinny być ze sobą utożsamiane.

Jedną z konsekwencji takiego ujęcia byłaby przy tym konieczność uznania, że pewne klasyczne argumenty na rzecz eliminatywizmu są nieskuteczne, ponieważ nie biorą pod uwagę wielopoziomowej organizacji systemów poznawczych oraz związanej z nią wielopoziomowości opisów i wyjaśnień tych systemów. Istnienie przekonań i pragnień nie zakłada, że musimy je znaleźć *wewnątrz* mechanizmów wyjaśniających fenomeny kognitywne. Jedynym poziomem organizacji, na którym należy szukać przekonań i pragnień, jest poziom całych systemów poznawczych (osób, organizmów). Trzeba przyznać, że idee tego rodzaju nie są całkowicie nowe w filozofii umysłu (por. Bennett, Hacker 2003, McDowell 1994, de Pinedo, Noble 2008, Rudder Baker 1995). Jednak ani ich znaczenie dla odpowiedniego zrozumienia problemu (problemów) reprezentacji, ani fakt, że naturalnie łączą się z koncepcją wyjaśnienia mechanistycznego, nie są na razie powszechnie rozpoznane i docenione.

3. WILLIAM RAMSEY I WYMÓG OPISU ZADAŃ

Oczywiście, uwagi te pozostawiają otwarte pytania o to, *czym* dokładnie miałyby być mechanizmy wykorzystujące reprezentacje umysłowe oraz *czy* takie mechanizmy są w ogóle możliwe.

Okazuje się, że bardzo istotne kroki w kierunku sformułowania warunków bycia mechanistycznym wyjaśnieniem reprezentacyjnym zostały poczynione przez Williama Ramseya w pracy *Representation Reconsidered* (Ramsey 2007). Chociaż nie przyjmuje *explicite* mechanistycznego podejścia do natury wyjaśniania w kognitywistyce, to wyniki jego rozważań są z mechanicyzmem zdecydowanie zbieżne. Zasadniczy cel teoretyczny Ramseya w *Representation Reconsidered* polega na (1) określeniu kryteriów, według których należy oceniać eksplanacyjną wartość pojęcia reprezentacji umysłowej, oraz (2) przeanalizowaniu według tych kryteriów rzeczywistej praktyki eksplanacyjnej kognitywistów, celem określenia, czy pojęcie reprezentacji rzeczywiście stanowi istotne narzędzie tej praktyki.

Ramsey wychodzi od przypuszczenia, że podkreślane ostatnio problemy reprezentacji nie wyrastają jedynie z powstania nowych naukowych koncepcji umysłu i procesów poznawczych, które obywają się bez postulowania wewnętrznych reprezentacji. Są również pochodną tego, że filozofowie umysłu oraz kognitywiści w za-

sadzie od czasu powstania kognitywistyki jako dyscypliny posługują się pojęciem reprezentacji umysłowej w sposób nazbyt swobodny, nazywając „reprezentacjami” struktury, które nie odgrywają w systemie poznawczym *roli bycia reprezentacją*, a w każdym razie nie pełnią jej w żadnym eksplanacyjnie wartościowym, informatywnym sensie tego terminu. Istnieje tendencja do nazywania „reprezentacją” dowolnej struktury neuronalnej bądź obliczeniowej pośredniczącej między percepcją a zachowaniem. Zasadniczym zamierzeniem *Representation Reconsidered* jest odróżnienie tych sposobów rozumienia reprezentacji w naukach kognitywnych, które przypisują określonym elementom systemu poznawczego funkcje rzeczywiście i nietrywialnie reprezentacyjne, od takich, przy których moglibyśmy obyć się bez posługiwania się terminem „reprezentacja” bez żadnych strat w rozumieniu i wyjaśnianiu badanych zjawisk.

Filtrem, który według Ramseya może umożliwić nam odróżnienie wyjaśnień prawomocnie i wartościowo reprezentacyjnych od tych, które są reprezentacyjne jedynie pozornie lub w jakimś nieciekawym i trywialnym sensie, jest trzystopniowa procedura, którą nazywa „wymogiem opisu zadań” (*job description challenge*).

Po pierwsze, polega ona na rekonstrukcji przednaukowego pojęcia reprezentacji¹⁰, koncentrując się na *rolach funkcjonalnych* przypisywanych reprezentacjom. Ramsey stwierdza, że pojęcie reprezentacji ma korzenie nie w teoriach naukowych, lecz w zdrowym rozsądku oraz koncepcjach filozoficznych (stanowiących do pewnego stopnia eksplikację pojęć zdroworozsądkowych). Postuluje, że przednaukowe pojęcie reprezentacji ma naturę prototypową. Różnego rodzaju obiekty klasyfikowane są jako reprezentacje nie dlatego, że spełniają konieczne i wystarczające warunki bycia reprezentacją, lecz dlatego, że posiadają odpowiednio wiele własności uznawanych za prototypowo reprezentacyjne. Ramsey twierdzi jednocześnie, że istnieją w istocie dwa pojęcia reprezentacji w takim przednaukowym sensie. Zgodnie z pierwszym z nich — mającym źródło w psychologii potocznej — reprezentacjami są postawy propozycjonalne (przekonania, pragnienia, nadzieje itp.), które charakteryzują się posiadaniem niewywiezionej treści intencjonalnej (tzn. posiadanie tej treści nie wymaga bycia interpretowanym lub interpretowalnym przez podmiot intencjonalny) oraz własności przyczynowych związanych z posiadaniem takiej, a nie innej treści. Zgodnie z drugim pojęciem reprezentacjami są różnego rodzaju obiekty lub zdarzenia o charakterze zewnętrznym, które wykorzystywane są przez podmiot intencjo-

¹⁰ Ramsey (2007) argumentuje za tym, że wszystkie próby ominięcia tego etapu i wprowadzenia technicznej definicji terminu „reprezentacja” skończyły się niepowodzeniem. Wszelkie tego rodzaju definicje są albo zbyt szerokie, albo zbyt wąskie. Według Ramseya te niepowodzenia nie wynikają po prostu z faktu, że proponowane do tej pory przez badaczy definicje okazały się za mało przemyślane. Głębszą przyczyną tego stanu rzeczy jest to, że pojęcie reprezentacji jest prototypowe, w związku z czym podanie listy warunków osobno koniecznych, a łącznie wystarczających do bycia reprezentacją jest niewykonalne. Właśnie dlatego Ramsey decyduje się na zaufanie (w pewnych granicach, zob. główny tekst artykułu) preteoretycznym intuicjom dotyczącym poprawnego posługiwania się terminem „reprezentacja”.

nalny do tego, by informowały go o innych obiektach czy zdarzeniach bądź zastępowały¹¹ mu inne obiekty lub zdarzenia. Przykładami reprezentacji w tym sensie mogą być termometry, liczniki prędkości, słoje drzewa, obrazy i mapy.

Po drugie, według Ramseya należy dokonać rekonstrukcji praktyki pojęciowej kognitywistów. Chodzi o dokładne wskazanie, jakie rodzaje stanów czy struktur nazywane są rutynowo reprezentacjami *zgodnie z rzeczywistą praktyką pojęciową badaczy*. Należy jednocześnie zwrócić szczególną uwagę na to, jakie rodzaje *funkcji* przysługują tym stanom czy strukturom, na czym polegają „zadania” pełnione przez nie w ramach systemu poznawczego. Ramsey wyróżnia łącznie pięć pojęć reprezentacji występujących w kognitywistyce: (1) potoczne pojęcie reprezentacji jako postaw propozycjonalnych, (2) pojęcie typu „wejście—wyjście” (IO-reprezentacje), (3) pojęcie reprezentacji strukturalnych czy też reprezentacji jako symulacji (S-reprezentacje), (4) pojęcie reprezentacji jako receptorów, (5) pojęcie reprezentacji ukrytych (niejawnych). Trzy pierwsze pojęcia są według Ramseya wykorzystywane głównie w ramach klasycznego już symbolicznego modelu obliczeniowej architektury umysłu. Dwa pozostałe rozpowszechniły się dopiero wraz z powstaniem koneksjonizmu oraz rozwojem neuronauki poznawczej.

Po trzecie wreszcie, należy powiązać dwa pierwsze kroki, zadając pytanie o to, które ze struktur przez kognitywistów nazywanych zwykle „reprezentacjami” odgrywają role funkcjonalne podobne do tych odgrywanych przez „reprezentacje” w przednaukowym znaczeniu tego słowa. To pytanie wyznacza istotę wymogu opisu

¹¹ Ujmowanie roli pełnionej przez reprezentacje za pomocą kategorii „zastępowania czegoś” bywa krytykowane (Sloman 2011, Miłkowski 2013). Dziękuję anonimowemu recenzentowi za zwrócenie uwagi na ten fakt. Przeciwnicy takiego ujmowania funkcji reprezentacji zauważają, że jeśli x ma (dosłownie) zastąpić y , to x musi być zdolne do odgrywania ról przyczynowych, które pełnione są na ogół przez y . Reprezentacje nie posiadają jednak takich samych ról przyczynowych jak to, co jest przez nie reprezentowane. Nie możemy przecież używać reprezentacji w ten sam sposób, w jaki używamy reprezentowanego obiektu. Mówiąc obrazowo, nie możemy obrać i zjeść *reprezentacji* banana (Sloman 2011; o ile rzecz jasna do reprezentowania jednego banana nie używamy innego banana). Czy można zatem opisać rolę reprezentacji jako „zastępowanie” tego, co reprezentowane? Wydaje się, że cały problem jest werbalny, a nie merytoryczny. Nikt nie twierdzi poważnie, że reprezentacja może *dosłownie* zastąpić to, co reprezentowane, na przykład tak, jak jeden nauczyciel matematyki może zastąpić innego. Kiedy mowa jest o „zastępowaniu” czegoś przez reprezentację, chodzi jedynie o to, że reprezentacja ta *zapośrednicza* kontakt poznawczy (lub praktyczny) z tym, co reprezentowane. Reprezentacje umożliwiają podejmowanie decyzji i wyciąganie wniosków na temat pewnego obiektu (stanu rzeczy, procesu itd.), a także antycypowanie przebiegu potencjalnych interakcji z tym obiektem bez konieczności wchodzenia z nim w bezpośredni kontakt. Możemy wyciągnąć wniosek o pożarze, patrząc na unoszący się dym, bez konieczności wchodzenia w kontakt z samym ogniem; korzystając z mapy możemy zaplanować trasę, unikając w ten sposób wypróbowywania różnych możliwych ścieżek „po omacku”. W takim właśnie sensie dym i mapa mogą (poznawczo) „zastąpić” to, co reprezentowane; mogą być wykorzystywane w określonych celach (poznawczych i praktycznych) „zamiast” samego reprezentowanego obiektu. Aby zastępować coś w takim sensie, nie trzeba posiadać tych samych własności przyczynowych co reprezentowany obiekt.

zadań. Jeśli chcemy uznać wyjaśnienie danego fenomenu kognitywnego za prawdopodobnie reprezentacyjne, to wewnętrzne struktury postulowane w tym wyjaśnieniu jako „reprezentacje” muszą w sposób naturalny i intuicyjny *służyć za reprezentacje*. To zaś ma miejsce, gdy ich role funkcjonalne są w pewnym stopniu analogiczne do tych pełnionych przez reprezentacje w sensie przednaukowym. Nie oznacza to oczywiście, że przednaukowe i naukowe pojęcia reprezentacji muszą być dokładnie takie same. Dla przykładu mapy i znaki językowe są reprezentacjami tylko o tyle, o ile służą za reprezentacje podmiotom intencjonalnym. Badacze-kognitywiści nie mogą — pod groźbą popełnienia błędu homunkularnego — postulować istnienia podmiotów intencjonalnych wewnątrz systemów poznawczych. Reprezentacje w sensie naukowym będą więc musiały różnić się jakoś od reprezentacji w sensie przednaukowym. Jednak aby spełnić wymóg opisu zadań — aby odgrywać rolę eksplanacyjną *jako reprezentacja* — opis funkcjonowania reprezentacji wewnętrznych (neuralnych lub obliczeniowych) musi w pewnym stopniu odpowiadać temu, jak funkcjonują reprezentacje w sensie przednaukowym. Tylko przyjmując przednaukowe rozumienie reprezentacji za punkt wyjścia, możemy uniknąć sytuacji, w której cokolwiek w arbitralny sposób nazwane może zostać „reprezentacją”. Według Ramseya stosowanie procedury opartej na wymogu opisu zadań pozwoli udzielić odpowiedzi — w odniesieniu do każdego pojęcia reprezentacji wymienionego w drugim kroku procedury — na pytanie o to, czy przyjęcie perspektywy reprezentacyjnej pozwala w danym przypadku zrozumieć jakieś zjawisko lepiej niż w sytuacji, w której porzucimy terminologię reprezentacjonistyczną; tzn. czy otrzymujemy dzięki pojęciu reprezentacji jakiś wymierny eksplanacyjny zysk, którego zostalibyśmy pozbawieni, rezygnując z myślenia w kategoriach reprezentacyjnych.

Raz jeszcze należy podkreślić, że sam Ramsey nie sytuuje swojego wyводу w kontekście koncepcji wyjaśniania mechanistycznego. Co więcej, mimo że wyjaśnianie odgrywa w jego pracy zasadniczą rolę, autor nie przyjmuje żadnych założeń dotyczących jego natury. Nie ulega jednak wątpliwości, że jego metodologia ocenia wartości eksplanacyjnej reprezentacji jest zgodna z mechanicyzmem. Jest tak dlatego, że idea „wymogu opisu zadań” oparta jest na założeniu, iż bycie reprezentacją w wartościowym eksplanacyjnie sensie zakłada *służenie za* lub *funkcjonowanie jako* reprezentacja w obrębie systemu poznawczego. Założenie to, jak już zauważyliśmy, naturalnie wynika z przyjęcia postawy mechanistycznej, która każe przecież rozpatrywać reprezentacje jako komponenty mechanizmów odpowiedzialne za operację reprezentowania (tj. funkcjonowania jako reprezentacja). Zatem jeśli dana struktura postulowana przez naukowców jako reprezentacja spełnia wymóg opisu zadań, to można o niej powiedzieć, że wykonuje operację polegającą na reprezentowaniu. Przyjmując założenia mechanicyzmu, możemy dodatkowo powiedzieć, że struktura ta zawdzięczać będzie swój status reprezentacji temu, jaką rolę odgrywa *względem innych komponentów szerszego mechanizmu*, którego komponent stanowi. Wymóg opisu zadań jest więc dla mechanicysty narzędziem przydatnym do przeprowadzenia linii demarkacyjnej między mechanizmami reprezentacyjnymi a niere-

prezentacyjnymi; albo też, inaczej, między mechanizmami reprezentacyjnymi w sensie nietrywialnym i eksplanacyjnie wartościowym a mechanizmami quasi-reprezentacyjnymi, tj. takimi, które nie zasługują na miano reprezentacyjnych¹².

Nie sposób opisać tu szczegółowo wniosków, jakie wynikają według Ramseya z zastosowania procedury opartej na wymogu opisu zadań. Poruszając się na poziomie najbardziej ogólnych konkluzji autora, można jednak powiedzieć, że są one w dużym stopniu nieprzychylnie reprezentacjonizmowi. Ramsey stawia między innymi tezę, że wbrew powszechnie stosowanej nomenklaturze reprezentacyjnej kognitywistyka już od dłuższego czasu — poprzedzającego nawet powstanie otwarcie rezygnujących z reprezentacji nurtów enaktywnych i ucieleśnionych — znajduje się w fazie antyreprezentacjonistycznej. Mimo to Ramsey twierdzi, że wśród wszystkich rozpowszechnionych wśród badaczy sposobów rozumienia terminu „reprezentacja” istnieją dwa, które są nietrywialnie reprezentacyjne. Dotyczy to pojęć reprezentacji umysłowych jako IO-reprezentacji oraz S-reprezentacji¹³. Innymi słowy, IO-reprezentacje oraz S-reprezentacje pełnią w ramach systemu poznawczego funkcje o charakterze reprezentacyjnym. Tak więc reprezentacyjne wyjaśnienia są możliwe — spełniają wymóg opisu zadań — o ile odwołują się do jednego z tych dwóch rodzajów reprezentacji.

Założmy na potrzeby tego wywodu, że negatywne diagnozy Ramseya w odniesieniu do wartości trzech pozostałych sposobów rozumienia reprezentacji przez kognitywistów są poprawne. Założmy też, że Ramsey ma rację, twierząc, iż stosowność pojęcia IO-reprezentacji ogranicza się do tradycyjnego, symbolicznego paradygmatu w sztucznej inteligencji. Choć z tezą tą bez wątplenia można się spierać, czysto prowizorycznie przyjmijmy także, że Ramsey ma rację, diagnozując ów paradygmat jako „nieaktualny” i należący (lub przechodzący) już do historii kognitywistyki. Jeśli wszystkie te przesłanki są prawdziwe, to jedynym wartościowym eksplanacyjnie dla kognitywistyki rodzajem reprezentacji umysłowych pozostają S-reprezentacje (por. przypis 13). Następny paragraf poświęcony jest próbie określenia, jak powinniśmy rozumieć mechanizmy S-reprezentacyjne.

¹² Zauważmy, że nie chodzi o to, że niereprezentacyjne lub quasi-reprezentacyjne wyjaśnienia mechanistyczne nie mogą być poprawne. W istocie może okazać się, że mechanizmy nieoparte na reprezentacjach często doskonale wyjaśniają interesujące badaczy zjawiska.

¹³ Konkluzja ta jest według Ramseya tym bardziej nieprzychylna dla reprezentacjonizmu, że przydatność IO- oraz S-reprezentacji ogranicza się do symbolicznego paradygmatu w sztucznej inteligencji. Teza ta nie wydaje się jednak poprawna w odniesieniu do S-reprezentacji i stanowi, być może, najsłabszy element pracy Ramseya. Do S-reprezentacji odwołuje się przecież także wiele koncepcji opartych na modelowaniu koneksjonistycznym czy teorii sterowania (zob. dalszą część artykułu oraz Calvo, Rodriguez 2009, Grush 2008, Shagrir 2012).

4. MECHANIZMY OPARTE NA S-REPREZENTACJACH

S-reprezentacje należą do klasy reprezentacji, w których relacja między samą reprezentacją (nośnikiem reprezentacji) a tym, co reprezentowane, zasada się na podobieństwie. Reprezentacjami opartymi na podobieństwie są między innymi mapy, pomniejszone modele przestrzenne (np. modele Układu Słonecznego), drzewa genealogiczne i fotografie. Wszystkie te reprezentacje pozwalają wyciągnąć wiele wniosków na temat tego, co reprezentują: struktury przestrzennej danego obszaru albo układu planetarnego, układu relacji pokrewieństwa w określonej rodzinie albo wyglądu osoby, którą przedstawia fotografia. Funkcję pełnią przez takie reprezentacje można opisać, stwierdzając, że reprezentacje te *zastępują* nam reprezentowany obiekt, proces czy zdarzenie. Odgrywają tę rolę właśnie dlatego, że przypominają pod pewnymi względami to, co reprezentowane. Mapa pozwala orientować się w terenie, ponieważ jej topograficzna struktura odzwierciedla czy też obrazuje topograficzną strukturę określonego (reprezentowanego) terenu; z kolei określone elementy mapy reprezentują określone elementy terenu (np. określone elementy graficzne reprezentują kościoły) dzięki temu, że te pierwsze zajmują w obrębie struktury topograficznej mapy jako całości miejsce odpowiadające umiejscowieniu tych drugich w obrębie reprezentowanego terenu.

Mówiąc o S-reprezentacjach, mamy na myśli wypadki, w których podobieństwo reprezentacji i tego, co reprezentowane, nie polega na dzieleniu określonych cech fizycznych (np. koloru), lecz na podobieństwie *strukturalnym* (por. O'Brien, Opie 2004, Ramsey 2007, Swoyer 1991). Z S-reprezentowaniem mamy do czynienia wtedy, gdy *układ* czy *wzór relacji* zachodzących między elementami reprezentacji przypomina układ czy wzór relacji zachodzących między elementami tego, co reprezentowane. Dwa złożone obiekty (zdarzenia, procesy) S oraz V są strukturalnie podobne, jeśli przynajmniej niektórym elementom S oraz relacjom zachodzącym między elementami S można przyporządkować elementy V i zachodzące między nimi relacje tak, że jeżeli określona relacja R zachodzi między elementami x i y w S , to odpowiadająca R relacja R' zachodzi między dwoma (odpowiadającymi x i y) elementami x' i y' w V (bardziej techniczne omówienia strukturalnego podobieństwa zawarte są m.in. w O'Brien, Opie 2004 i Swoyer 2001). W niektórych wypadkach struktura reprezentacji i tego, co reprezentowane, jest tego samego rodzaju; np. przestrzenna struktura mapy reprezentuje przestrzenną strukturę terenu. W innych wypadkach rodzaje struktur są jednak różne; np. w (poprawnie reprezentującym) drzewie genealogicznym układ przestrzenny strzałek i napisów odzwierciedla układ relacji pokrewieństwa między członkami rodziny.

Rzecz jasna zachodzenie strukturalnego podobieństwa nie czyni jeszcze czegoś S-reprezentacją; jest ono konieczne do bycia S-reprezentacją, lecz niewystarczające. Wszystkie wymienione do tej pory rodzaje S-reprezentacji zakładają istnienie podmiotu interpretującego bądź stosującego je *jako reprezentacje*. Mapy, drzewa genealogiczne, modele przestrzenne i fotografie reprezentują tylko o tyle, o ile wykorzy-

stywane są jako reprezentacje przez podmioty intencjonalne. Tym samym są reprezentacjami nie tylko dzięki zachodzeniu strukturalnego podobieństwa, lecz także dzięki pełnieniu odpowiedniej funkcji. Fakt ten ma zasadnicze znaczenie, jeśli zajmujemy się problemem tego, czy S-reprezentacje mogą stanowić element mechanistycznych wyjaśnień formułowanych przez przedstawicieli nauk kognitywnych. Jeśli S-reprezentacje mają odgrywać rolę eksplanacyjną w kognitywistyce, to muszą stanowić funkcjonalne komponenty wewnętrznych, neurobiologicznych lub obliczeniowych mechanizmów. Aby tak było, wewnętrzne S-reprezentacje powinny służyć za reprezentacje w obrębie tego mechanizmu, i to w sposób na tyle zbliżony do tego, jak funkcjonują mapy czy drzewa genealogiczne, aby mówienie o reprezentowaniu było uzasadnione z punktu widzenia eksplanacyjnych zamysłów naukowców (tj. aby wewnętrzne S-reprezentacje czyniły zadość wymogowi opisu zadań). Co jednak oczywiste, opisując działanie mechanizmów stojących u podstaw zdolności kognitywnych, nie możemy postulować istnienia interpretujących podmiotów gdzieś w centralnym układzie nerwowym. Czy możemy nadać sens idei, że mózg coś S-reprezentuje — że istnieje w nim coś, co pełni rolę „wewnętrznej mapy” — nie popełniając jednocześnie błędu homunkularnego? Jeśli odpowiedź na to pytanie jest twierdząca, to okaże się, że pojęcie S-reprezentacji można „zmechanizować”.

Na tym etapie warto powrócić do pracy Ramseya (2007). Jak już zostało zaznaczone, przekonuje on, że pojęcie S-reprezentacji spełnia wymóg opisu zadań. W kontekście przyjętych tu założeń oznacza to jednocześnie, że pojęcie to może stać się elementem wyjaśnień mechanistycznych. Procedura, którą przyjmuje Ramsey, argumentując za wartością eksplanacyjną S-reprezentacji, jest następująca: po pierwsze, każe on wyobrazić sobie niekontrowersyjny przykład S-reprezentacji, takiej, która funkcjonuje jako reprezentacja, ponieważ jest w odpowiedni sposób wykorzystywana przez podmiot intencjonalny. Po drugie, modyfikuje wyjściowy przykład w ten sposób, że to, co w pierwszym przykładzie było S-reprezentacją, nadal pełni maksymalnie podobną funkcję, jednak nie jest wykorzystywane przez żaden podmiot. Po trzecie pokazuje, że w zmodyfikowanym przykładzie nadal mamy do czynienia z S-reprezentacją w wartościowym (czy wręcz niezbędnym) eksplanacyjnie sensie, tj. że potrzebujemy pojęcia reprezentacji, aby zrozumieć czy wyjaśnić, co dzieje się w zmodyfikowanym przypadku.

Weźmy pod uwagę następującą, bardzo prostą ilustrację zastosowania opisanej wyżej strategii (Ramsey 2007: 196-199)¹⁴. Wyobraźmy sobie samochód jadący torem, którego boczne granice wyznaczone są przez wysoki mur. Chcąc zrozumieć, jak to się dzieje, że samochód skutecznie porusza się w obrębie toru — nie odbija się od krawędzi do krawędzi, lecz jedzie środkiem, zakręcając w odpowiednich momentach,

¹⁴ Przykład został tu skrócony. W oryginalnej wersji służy nie tylko do pokazania, że S-reprezentacje czynią zadość wymogowi opisu zadań, lecz także do wykazania, że tego samego nie możemy powiedzieć o reprezentacjach-receptorach, które miałyby reprezentować ze względu na relacje przyczynowo-nomologiczne; Ramsey 2007: 196-199.

pod odpowiednim kątem i w odpowiednim kierunku — odkrywamy, że wewnątrz znajduje się kierowca. Okazuje się jednak, że okna samochodu są nieprzezroczyste, a kierowca odpowiednio steruje samochodem dzięki temu, że dysponuje mapą odzwierciedlającą kształt toru. Jest to mapa elektroniczna, samochód oznaczony jest na niej czerwoną kropką, której położenie na mapie odpowiada każdorazowo położeniu samochodu w obrębie toru. Tym samym zyskujemy S-reprezentacyjne, lecz *nie* mechanistyczne wyjaśnienie zachowania samochodu. Zmodyfikujemy teraz ten przykład. Wyobraźmy sobie samochód skutecznie poruszający się po torze, ale pozbawiony kierowcy. Zamiast siedzącego w środku człowieka tym razem odkrywamy, że we wnętrzu samochodu znajduje się powierzchnia, w której wydrążony jest rowek o kształcie identycznym z kształtem toru, po którym porusza się samochód. Do rowka podłączony jest ster. Porusza się on wewnątrz rowka jednocześnie z pojazdem poruszającym się w obrębie toru, tak że ster każdorazowo znajduje się w miejscu rowka, które odpowiada umiejscowieniu samochodu na torze. Ster jest podłączony do kierownicy w ten sposób, że kierunek obrotu kierownicy — a przez to kierunek poruszania się samochodu — systematycznie odpowiada zmianom orientacji steru.

Czy zmodyfikowany, pozbawiony kierowcy samochód nadal stanowi system, którego działanie powinniśmy wyjaśniać w kategoriach reprezentacyjnych? Wydaje się, że odpowiedź jest tu twierdząca. Sposób funkcjonowania rowka, do którego podłączony jest ster, przypomina sposób, w jaki funkcjonowała mapa w samochodzie kierowanym przez człowieka. Satysfakcjonujące wyjaśnienie sukcesu nawigacyjnego pojazdu poruszającego się po torze wymaga powołania się na fakt, że rowek pełnił w mechanizmie odpowiadającym za ten sukces funkcję wewnętrznej mapy. Nie zrozumielibyśmy zachowania pozbawionego kierowcy samochodu bez powołania się na fakt, że strukturalno-przestrzenne własności tej mapy odpowiadały strukturalno-przestrzennym własnościom toru. Oczywiście, samo zachodzenie tego podobieństwa nie wyjaśnia działania pojazdu. Co ciekawe, sam Ramsey pisze o własnym przykładzie tak, jak gdyby to *samo podobieństwo* było wystarczającą podstawą do stwierdzenia, że pojazd jest system (S-)reprezentacyjnym¹⁵. Tymczasem podstawową rolę odgrywa tu *zarówno* zachodzenie odpowiedniości między rowkiem a torem, jak i *funkcjonalne położenie* rowka (wewnętrznej mapy) w obrębie szerszego mechanizmu wyjaśniającego sukces nawigacyjny samochodu.

Przydatnym — a w ogóle niewykorzystywanym przez samego Ramseya — pojęciem ujmującym funkcjonalne położenie rowka w obrębie mechanizmu jest pojęcie „konsumenta” reprezentacji. Zostało ono zaczerpnięte z prac Ruth Millikan (1984, 2002). Różnica polega jednak na tym, że Millikan wykorzystywała je, aby rozwiązywać problemy dotyczące treści reprezentacji (problem błędnej reprezentacji), natomiast tutaj służy nowemu filozoficznemu celowi, polegającemu na dookreśleniu funkcjonalno-mechanistycznych warunków bycia reprezentacją. Na obecne potrzeby

¹⁵ Naraża się on tym samym na proste zarzuty powołujące się na fakt, że podobieństwo strukturalne nie wystarczy do bycia reprezentacją (zob. np. Calvo, Rodriguez 2009).

możemy zdefiniować konsumenta reprezentacji jako komponent lub grupę komponentów mechanizmu, które są funkcjonalnie zależne od innego komponentu (reprezentacji) w ten sposób, że poprawne wykonywanie przez nie określonych operacji jest systematycznie, nieprzypadkowo zależne od tego, czy ten drugi komponent (reprezentacja) pozostaje w stosunku strukturalnego podobieństwa do określonych okoliczności środowiskowych¹⁶. Innymi słowy, konsumenci *używają* pewnego komponentu mechanizmu jako reprezentacji albo też komponent ten *służy* im za reprezentację. W przykładzie Ramseya konsumentem reprezentacji byłby ster, a mówiąc szerzej, cały podsystem sterujący położeniem kół pojazdu, złożony ze steru i kierownicy.

Potraktujmy zatem mechanizm odpowiadający za skuteczne poruszanie się samochodu w opisanym przykładzie jako klarowny, choć prosty przykład mechanizmu S-reprezentacyjnego. Mechanizm ten jest reprezentacyjny dlatego, że jeden z jego komponentów służy za reprezentację innym komponentom w jego obrębie. Spróbujmy uogólnić ten przykład i sformułować na jego podstawie listę warunków bycia mechanizmem S-reprezentacyjnym:

(i) W ramach mechanizmu istnieje komponent, którego poprawne funkcjonowanie zależy od tego, czy komponent ten znajduje się w relacji podobieństwa strukturalnego z okolicznościami środowiskowymi. Komponent ten możemy nazwać „reprezentacją” lub „nośnikiem reprezentacji” (w przypadku samochodu z opisanego przykładu jest to wewnętrzna mapa toru).

(ii) Działanie (funkcjonowanie) co najmniej jednego spośród pozostałych komponentów mechanizmu jest w systematyczny i nieprzypadkowy sposób zależne od stanu, w którym znajduje się nośnik reprezentacji. Komponenty tego rodzaju nazwać możemy konsumentami reprezentacji (w wypadku samochodu jest to podsystem sterujący, który jest podłączony do mapy tak, że jego działanie jest systematycznie zależne od kształtu mapy).

(iii) Poprawne funkcjonowanie konsumenta reprezentacji zależy od tego, jakie zachodzą (lub będą zachodzić) okoliczności środowiskowe, jednak nie istnieje bezpośrednia przyczynowa zależność między działaniem konsumentów a warunkami środowiskowymi (np. poprawne prowadzenie samochodu przez podsystem sterujący zależy od tego, jaki jest kształt toru, ale tor nie oddziałuje bezpośrednio na sposób działania tego podsystemu).

(iv) Zależność wymieniona w (ii) charakteryzuje się tym, że konsument poprawnie funkcjonuje w ramach mechanizmu tylko wtedy, gdy zachodzi strukturalne podobieństwo między nośnikiem reprezentacji a warunkami środowiskowymi. Tym samym możemy powiedzieć, że nośnik reprezentacji „zastępuje” konsumentowi warunki środowiskowe (w wypadku samochodu podsystem sterujący jest połączony z wewnętrzną mapą w taki sposób, że poprawnie kieruje on samochodem tylko wtedy, gdy kształt mapy jest strukturalnie podobny do kształtu toru).

¹⁶ Dokładniej — tych aspektów (obiektów, zdarzeń, procesów), które są istotne dla sposobu działania konsumentów, a do których konsumenci nie mają bezpośredniego dostępu przyczynowego.

(v) Działanie takiego mechanizmu — obejmującego nośnik reprezentacji, jego konsumenta (konsumentów) oraz ewentualnie pozostałe, zorganizowane w określony sposób komponenty — wyjaśnia jakieś zjawisko (zjawiska) kognitywne lub zdolność do skutecznego działania w środowisku (np. mechanizm złożony z wewnętrznej mapy i podsystemu sterującego wyjaśnia skuteczne pokonywanie toru przez samochód z podanego przykładu).

Rzecz jasna wszystkie poczynione do tej pory uwagi służą pokazaniu, że formułowanie pełnoprawnie reprezentacyjnych wyjaśnień mechanistycznych jest możliwe. Uproszczone przykłady z samochodami jednak niewiele nam mówią o tym, jak z określonymi problemami poznawczymi radzą sobie realne, biologiczne mózgi. Reprezentacjonizm (przynajmniej w odniesieniu do określonych typów zjawisk poznawczych) *może* być podejściem poprawnym w świetle koncepcji wyjaśniania mechanistycznego, ale to, *czy jest* on poprawny, pozostaje kwestią empiryczną. To drugie zagadnienie wymagałoby oczywiście osobnego potraktowania. Warto jednak przynajmniej pobieżnie zwrócić uwagę na fakt, że pojęcie S-reprezentacji jest w istocie od dawna stosowane przez kognitywistów do wyjaśniania niektórych interesujących ich zjawisk. Do reprezentacji opartych na podobieństwie strukturalnym odwołują się niektóre klasyczne już koncepcje z zakresu psychologii (por. Johnson-Laird 1983) oraz sztucznej inteligencji (por. Laird, Newell, Rosenbloom 1987). Także stosunkowo dobrze ugruntowana empirycznie teoria *mind-reading* („czytania w myślach”) autorstwa Alvina Goldmana (2006) zasadza się na tezie, że zdolność do przypisywania innym osobom stanów umysłowych oparta jest na S-reprezentowaniu (symulowaniu) tych stanów. Również nowe podejścia w zakresie tworzenia neurobiologicznie realistycznych obliczeniowych modeli działania mózgu odwołują się do S-reprezentacji (por. np. Eliasmith 2003, O’Brien, Opie 2004, Shagrir 2012).

W tej grupie zapewne najciekawsza jest koncepcja reprezentacji jako emulacji stworzona przez Ricka Grusha (1997, 2004). Zasadza się ona na tezie, że system motoryczny korzysta z wewnętrznych modeli systemu szkieletowo-mięśniowego oraz środowiska, które pozwalają mu „przewidzieć” sensoryczne konsekwencje inicjowanych działań, zanim jeszcze dotrą do niego stosowne sygnały zwrotne z modalności zmysłowych. Upraszczając, emulatory stanowiłyby w takiej koncepcji reprezentacje lub nośniki reprezentacji, których konsumentem byłby system motoryczny. Ciekawym aspektem teorii Grusha (oraz teorii do niej podobnych) jest fakt, że jest ona zarówno nietrywialnie reprezentacyjna, jak i zgodna z duchem nowoczesnego, ucieleśnionego podejścia do natury umysłu i poznania, w ramach którego traktuje się procesy poznawcze jako zakorzenione w procesach sensomotorycznych. Jeśli poprawna jest teza, że koncepcję emulacji można rozszerzyć tak, by wyjaśniała również wyższe procesy poznawcze związane z posługiwaniem się pojęciami (zob. Pezzulo 2011), to stanowiłoby to bardzo dobrą wiadomość dla każdego reprezentacjonisty.

PODSUMOWANIE

Spór o eksplanacyjny status reprezentacji umysłowych w kognitywistyce prowadzony jest często bez przyjmowania w sposób otwarty określonych założeń dotyczących tego, na czym polega wyjaśnianie w kognitywistyce *jako takie*. Celem tego artykułu było umieszczenie sporu o reprezentacjonizm w kontekście koncepcji, zgodnie z którą wyjaśnianie zjawisk przez kognitywistów polega na odkrywaniu mechanizmów tych zjawisk. Pokazane zostało, że przyjęcie takiej perspektywy pozwala przeformułować niektóre problemy związane z reprezentacjami umysłowymi (np. odróżnić (anty)reprezentacjonizm lokalny od globalnego) oraz inaczej rozłożyć akcenty, jeśli chodzi o wagę poszczególnych zagadnień (np. dowartościować problem funkcjonalnych kryteriów bycia reprezentacją, częściowo uniezależniając go od problemu naturalizacji treści intencjonalnej). Jednocześnie podjęta tu została próba określenia warunków, w których wyjaśnienie mechanistyczne może zostać uznane za reprezentacyjne. Oznaczało to określenie funkcjonalnych kryteriów bycia reprezentacją umysłową czy też służenia za reprezentację umysłową w obrębie mechanizmu. Jak się okazało, koncepcja wyjaśniania mechanistycznego jest do pogodzenia z reprezentacjonizmem, przynajmniej dopóki postulowane reprezentacje są S-reprezentacjami, czyli takimi, których związek z tym, co reprezentowane, oparty jest na podobieństwie strukturalnym. Mówiąc precyzyjniej, zgodnie z poczynionymi tu ustaleniami mechanizm reprezentacyjny to taki, w którym (1) rola jednego z komponentów (reprezentacji czy nośnika reprezentacji) polega na pozostawaniu w relacji podobieństwa strukturalnego do określonych okoliczności czy warunków środowiskowych, ponieważ (2) co najmniej jeden inny komponent mechanizmu pełni rolę konsumenta reprezentacji, tj. jego poprawne działanie jest funkcjonalnie uzależnione od tego, czy reprezentacja jest strukturalnie podobna do tego, co reprezentowane¹⁷.

¹⁷ Przedstawione w tym artykule tezy są w dużym stopniu zbieżne z poglądami bronionymi przez Marcina Miłkowskiego (por. Miłkowski 2009, 2013). Podobieństwa dotyczą przede wszystkim wspólnych punktów wyjścia oraz założeń metodologicznych (co ważne, uwaga ta nie zawęży się do stanu *polskiej* twórczości filozoficznej, ponieważ systematyczne łączenie mechanicyzmu z zagadnieniem reprezentacji jest stosunkowo nową i rzadko omawianą ideą także w skali światowej). Po pierwsze, oba stanowiska (tj. wyrażone w tym tekście oraz to bronione przez Miłkowskiego) wychodzą od inspirowanych mechanicyzmem założeń dotyczących natury wyjaśniania w kognitywistyce. Po drugie, w obu podobnie postrzega się konsekwencje przyjęcia perspektywy mechanistycznej dla problemu reprezentacji, przy czym konsekwencją nadrzędną jest konieczność wskazania, jakiego rodzaju własności funkcjonalne gwarantują mechanizmowi lub komponentowi mechanizmu status „reprezentacyjny”. Po trzecie, w obrębie obu stanowisk bardziej lub mniej otwarcie rozpoznaje się odrębność problemu eksplanacyjnego statusu reprezentacji od innych problemów związanych z pojęciem reprezentacji omawianych w filozofii umysłu (np. problemu naturalizacji treści umysłowej). Analogie kończą się jednak w momencie, gdy chodzi o pozytywne rozstrzygnięcia dotyczące natury reprezentacyjnych mechanizmów i reprezentacyjnego wyjaśniania. Po pierwsze, w obrębie przedstawianego tu stanowiska istotną rolę gra idea zachodzenia podobieństwa strukturalnego między nośnikiem reprezentacji a tym, co jest reprezentowane. Przyjmuje się tu też

Przedstawiony tu wywód, co warte podkreślenia, nie wyklucza istnienia pełnoprawnie reprezentacyjnych wyjaśnień mechanistycznych, które jednak odwoływałyby się do reprezentacji jakiegoś innego rodzaju, niż te oparte na podobieństwie strukturalnym. Kwestia ta pozostaje otwarta. Ramsey (2007) przedstawia wiele argumentów za tym, że inne rodzaje struktur nazywane powszechnie przez kognitywistów „reprezentacjami” — m.in. receptory albo tzw. reprezentacje ukryte — nie pełnią w istocie żadnych funkcji, które pozwalałyby uznać je za reprezentacje w sposób uzasadniony i eksplanacyjnie wartościowy. Nie ulega mimo to wątpliwości, że w świetle przeprowadzonego tu wywodu okazuje się, iż pełnoprawnie reprezentacyjne wyjaśnienia mechanistyczne są możliwe. Co więcej, wydaje się, że na takich wyjaśnieniach opiera się wiele wartościowych, dobrze udokumentowanych oraz, co ważne, nowych lub stosunkowo nowych koncepcji z zakresu nauk kognitywnych. Można więc uznać, że pozycja reprezentacjonizmu na obecnym etapie rozwoju kognitywistyki nie jest tak zła, jak to często przedstawiają przedstawiciele (radzykalnie) ucieleśnionych i enaktywnych podejść do badania umysłu i poznania.

BIBLIOGRAFIA

- Bechtel W. (1994), *Levels of Description and Explanation in Cognitive Science*, „Minds and Machines” 4(1), 1-25.
- Bechtel W. (2008), *Mental Mechanisms. Philosophical Perspectives on Cognitive Neuroscience*, London: Routledge.
- Bechtel W., Abrahamsen A. (1993), *Connectionism and the Future of Folk Psychology* [w:] *Natural and Artificial Minds*, R. Burton (red.), Albany (NY): SUNY Press.

za Ramseyem (2007), że w rozważaniach nad reprezentacjami ważną rolę odgrywa rozróżnienie między reprezentacjami opartymi na podobieństwie a wskaźnikowymi bądź receptorowymi reprezentacjami opartymi na kowariancji (według Ramsey’a te drugie nie spełniają wymogów nałożonych na wartościową eksplanacyjnie koncepcję reprezentacji). Obydwa te założenia są odrzucane przez Miłkowskiego. W jego teorii istotną funkcję pełni pojęcie informacji, a w przyjętym przez niego rozumieniu informacja zawsze jest pewnego rodzaju strukturą. Dodawanie zastrzeżeń o podobieństwie strukturalnym między nośnikiem a przedmiotem reprezentacji jest według niego teoretycznie zbędne. Miłkowski uznaje też, że w przyjętej przez niego perspektywie nie da się przeprowadzić rozróżnienia między reprezentacjami wskaźnikowymi i reprezentacjami opartymi na podobieństwie. Po drugie, Miłkowski twierdzi, że reprezentacje muszą być „konsumowane” nie przez jeden (jak w przedstawionej tu teorii), a dwa inne komponenty mechanizmu. Postuluje on przede wszystkim, że reprezentacje spełniają swoje role eksplanacyjne jedynie wtedy, gdy istnieje komponent mechanizmu zdolny do rozpoznawania błędów reprezentacyjnego. Prezentowana w tym artykule koncepcja nie zakłada takiego ograniczenia. Po trzecie, konsekwencją teorii Miłkowskiego jest ściśle powiązanie reprezentacji z obliczeniową naturą systemu poznawczego. Mówiąc precyzyjniej, obliczeniowość systemu jest dla Miłkowskiego warunkiem koniecznym zaistnienia w nim reprezentacji środowiska. Tymczasem stanowisko przedstawione w tym artykule nie rozstrzyga kwestii relacji między obliczeniowością systemu poznawczego a reprezentacjami. Pozostawia ono możliwość istnienia nieobliczeniowych mechanizmów reprezentacyjnych.

- Bechtel W., Abrahamsen A. (2005), *Explanation. A Mechanist Alternative*, „Studies in History and Philosophy of Science Part C” 36(2), 421-441.
- Bechtel W., Abrahamsen A. (2010), *Dynamic Mechanistic Explanation. Computational Modeling of Circadian Rhythms as an Exemplar for Cognitive Science*, „Studies in History and Philosophy of Science Part A” 41(3), 321-333.
- Bechtel W., Richardson R. C. (1993), *Discovering Complexity. Decomposition and Localization as Strategies in Scientific Research*, Princeton: Princeton University Press.
- Bennett M. R., Hacker P. M. S. (2003), *Philosophical Foundations of Neuroscience*, Malden (MA): Blackwell.
- Calvo P., Rodriguez A. G. (2009), *Where is Cognitive Science Heading?*, „Minds and Machines” 19(3), 301-318.
- Chemero A. (2009), *Radical Embodied Cognitive Science*, Cambridge (MA): MIT Press.
- Clark A., Toribio J. (1994), *Doing without Representing?*, „Synthese” 101(3), 401-431.
- Craver C. F. (2007), *Explaining the Brain. Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Clarendon Press.
- Craver C.F. (2008), *Functions and Mechanisms in Neuroscience* [w:] *Des neurones é la conscience. Neurophilosophie et philosophie des neurosciences*, P. Poirier, L. Faucher, E. Racine, E. Ennan (red.), Bruxelles: De Boeck Université.
- Craver C. F., Bechtel W. (2006), *Top-Down Causation without Top-Down Causes*, „Biology and Philosophy” 22(4), 547-563.
- Cummins R. (1975), *Functional Analysis*, „Journal of Philosophy” 72, 741-765.
- Cummins R. (1989), *Meaning and Mental Representation*, Cambridge (MA): MIT Press.
- Cummins R. (2000), *'How Does it Work?' versus 'What Are the Laws?'. Two Conceptions of Psychological Explanation* [w:] *Explanation and Cognition*, F. Keil, R. A. Wilson (red.), Cambridge (MA): MIT Press.
- Dennett D. (1995), *Osobowy i subosobowy poziom wyjaśniania. Ból* [w:] *Filozofia umysłu*, B. Chwedeńczuk (red.), Warszawa: Aletheia.
- De Pinedo M., Noble J. (2008), *Beyond Persons. Extending the Personal/Subpersonal Distinction to Non-Rational and Artificial Agents*, „Biology and Philosophy” 23(1), 87-100.
- Dretske F. (1988), *Explaining Behavior. Reasons in the World of Causes*, Cambridge (MA): MIT Press.
- Eliasmith C. (2003), *Moving beyond Metaphors. Understanding the Brain for What It Is*, „Journal of Philosophy” 100, 493-520.
- Fodor J. (1987), *Psychosemantics*, Cambridge (MA): MIT Press.
- Goldman A. (2006), *Simulating Minds. The Philosophy, Psychology and Neuroscience of Mindreading*, Oxford: Oxford University Press.
- Glennan S. (2002), *Rethinking Mechanistic Explanation*, „Philosophy of Science” 3(69): S342-S353.
- Gładziejewski P. (2012), *William Ramsey o psychologii potocznej, racjonalności i pojęciu reprezentacji w naukach kognitywnych*, „Diametros” 31, 33-55.
- Godfrey-Smith P. (2004), *On Folk Psychology and Mental Representation* [w:] *Representation in Mind. New Approaches to Mental Representation*, H. Clapin, P. Staines, P. Slezak (red.), Amsterdam: Elsevier, 147-162.
- Grush R. (1997), *The Architecture of Representation*, „Philosophical Psychology” 10(1), 5-23.
- Grush R. (2004), *The Emulation Theory of Representation. Motor Control, Imagery and Perception*, „Behavioral and Brain Sciences” 27(3), 377-442.
- Grush R. (2008), *Review of "Representation Reconsidered" by W. Ramsey*, „Notre Dame Philosophical Reviews”; <http://ndpr.nd.edu/news/23327/?id=12243>.

- Johnson-Laird P. (1983), *Mental Models. Towards a Cognitive Science of Language, Inference and Consciousness*, Cambridge (MA): Harvard University Press.
- Kaplan D. M., Bechtel W. (2009), *Dynamical Models. An Alternative or Complement to Mechanistic Explanations?*, „Topics in Cognitive Science” 3(2), 438-444.
- Laird J. E., Newell A., Rosenbloom P. S. (1987), *SOAR. An Architecture for General Intelligence*, „Artificial Intelligence” 33(1), 1-64.
- Machamer P., Darden L., Craver C. F. (2011), *Myślenie w kategoriach mechanizmów*, „Przegląd Filozoficzno-Literacki: Filozofia Biologii” 2-3(31), 175-190.
- McDowell J. (1994), *The Content of Perceptual Experience*, „Philosophical Quarterly” 44(175), 190-205.
- Millikan R. G. (1984), *Language, Thought and Other Biological Categories*, Cambridge: Cambridge University Press.
- Millikan R. G. (2002), *Varieties of Meaning*, Cambridge (MA): MIT Press.
- Miłkowski M. (2009), *Reprezentacje w systemach klasycznych i koneksjonistycznych*, „Studia z Kognitywistyki i Filozofii Umysłu” 3, 29-42.
- Miłkowski M. (2013), *Explaining the Computational Mind*, Cambridge (MA): MIT Press.
- O'Brien G., Opie J. (2004), *Notes toward a Structuralist Theory of Mental Representation* [w:] *Representation in Mind. New Approaches to Mental Representation*, H. Clapin, P. Staines, P. Slezak (red.), Amsterdam: Elsevier, 1-20.
- Pezzulo G. (2011), *Grounding Procedural and Declarative Knowledge in Sensorimotor Anticipation*, „Mind and Language” 24(5), 523-553.
- Piccinini G. (2007), *Computational Explanation and Mechanistic Explanation of Mind* [w:] *Cartographies of the Mind. The Interface between Philosophy and Cognitive Science*, M. de Caro, F. Ferretti, M. Marraffa (red.), Dordrecht: Springer, 23-36.
- Piccinini G., Craver C. F. (2011), *Integrating Psychology and Neuroscience. Functional Analyses as Mechanism Sketches*, „Synthese” 3 (183), 283-311.
- Ramsey W. (2007), *Representation Reconsidered*, Cambridge: Cambridge University Press.
- Ramsey W., Stich S., Garon J. (1990), *Connectionism, Eliminativism and the Future of Folk Psychology*, „Philosophical Perspectives” 4, 499-533.
- Rudder Baker L. (1994), *Explaining Attitudes. A Practical Approach to the Mind*, Cambridge: Cambridge University Press.
- Shagrir O. (2012), *Structural Representations and the Brain*, „British Journal for the Philosophy of Science” 63(3), 519-545.
- Slovan A. (2011), *Comments on “The Emulating Interview... with Rick Grush”*, „Avant” 2, 35-44.
- Van Gelder T. (1995), *What Might Cognition Be, If Not Computation?*, „Journal of Philosophy” 92(7), 345-381.
- Wimsatt W. (1974), *Reductive Explanation. A Functional Account*, „PSA. Proceedings of the Biennial Meeting of the Philosophy of Science Association”, 671-710.