

WPROWADZENIE

Pytanie o naturę i rolę reprezentacji umysłowych należy do najważniejszych zagadnień filozofii umysłu, a także psychologii i szeroko pojmowanych nauk o poznaniu. Pojęcie reprezentacji umysłowej jest stosowane powszechnie i pełni rozmaite funkcje. Odwołujemy się do reprezentacji zarówno w praktyce potocznej, gdy mówimy o tym, że ktoś myśli lub podejmuje jakieś decyzje, jak i w badaniach naukowych, gdy badamy zachodzące w mózgu i ciele procesy poznawcze oraz ich wpływ na czyjeś zachowanie.

We współczesnych rozważaniach filozoficznych nad naturą reprezentacji podstawowymi pytaniami są: Czym są reprezentacje?, Czym jest treść reprezentacji? oraz: Na mocy czego reprezentacje mają odniesienie i treść? Drugie pytanie wiąże się z naturą intencjonalności, czyli „byciem o”. „Bycie o” jest podstawową cechą reprezentacji. Reprezentacje mogą być o jakimś stanie rzeczy. Mogą być o przedmiotach rzeczywistych, a także o przedmiotach nierzeczywistych. I tak mogę myśleć np. zarówno o tej książce, jak i o przyszłych wakacjach. Powiemy, że reprezentacje mają treść, gdyż ten przedmiot jakoś opisują, mają także odniesienie, gdyż odnoszą się do przedmiotu, jeśli jest on rzeczywisty.

Intencjonalność wiąże się z adekwatnością reprezentacji. Reprezentacje mogą opisywać swoje przedmioty adekwatnie bądź nie. I tak moja myśl o tym, że dziś świeci słońce, jest trafna, gdy słońce wpada przez okno, natomiast myśl, że dziś jest piątek, jest błędna, gdy jest czwartek. Ponieważ reprezentacje mogą być poprawne lub błędne, pojęcie reprezentacji jest normatywne.

Założenie, że reprezentacje są normatywne, mogłoby być podważone przez odwołanie do prostoty: nienormatywne pojęcie reprezentacji byłoby prostsze. Ma ono jednak istotną wadę, gdyż zmusza do przyjęcia założenia, że wszystkie reprezentacje są z konieczności zawsze trafne bądź nie są reprezentacjami. To ma jednak bardzo negatywne konsekwencje. Pojęcie reprezentacji jest związane z racjonalnością instru-

mentalną, gdy jest stosowane do opisu ludzkich i zwierzęcych zachowań w kategoriach intencjonalnych. Racjonalność instrumentalna polega na doborze właściwych środków do założonego wcześniej celu (Weber 2002).

Wyobraźmy sobie Jasia, który chce dojechać autobusem do szkoły, i rozważmy dwa przypadki. W pierwszym Jaś postępuje zgodnie z rozkładem jazdy, aby dojechać do szkoły. Wówczas dobiera właściwe środki do zamierzonego celu. W drugim Jaś pamięta nieaktualny rozkład jazdy. Nie wie, że zgodnie z aktualnym rozkładem autobus nie przyjeżdża już o tej porze i dlatego nadal czeka na autobus.

Bez pojęcia reprezentacji umysłowej – i to błędnej, której paradygmatycznym przykładem może być fałszywe przeświadczenie Jasia, że autobus przyjedzie o 7:45 – wyjaśnienie zachowania studenta jako zachowania racjonalnego byłoby bardzo trudne. Przecież Jaś podejmuje działanie, które nie doprowadza go do zamierzonego celu: czeka, choć autobus nie przyjedzie. Wyrugowanie możliwości posiadania reprezentacji, w tym reprezentacji błędnych, prowadziłoby więc do niepożądanego konsekwencji, że każde obiektywnie nietrafne działanie byłoby oznaką irracjonalności. Ponieważ błędy w działaniu są powszechne, prowadziłoby to do powszechnego przypisywania irracjonalności¹. Tymczasem możemy zachować pojęcie racjonalności w wypadku pomyłek, gdy dysponujemy pojęciem błędnej reprezentacji. Jaś, który ma błędną reprezentację godziny przyjazdu autobusu, zachowuje się racjonalnie, ponieważ ma powody, by sądzić, że autobus przyjedzie o danej porze.

Analogiczne rozumowanie można przeprowadzić w odniesieniu do zwierząt. Edward Tolman wykonał eksperyment, w którym szczur nauczył się układu ramion labiryntu, mimo że nie było tam jedzenia mogącego stanowić dla zwierzęcia nagrodę. Kiedy jednak w labiryncie znalazł się pokarm, szczur natychmiast znajdował go, mimo że nie był głodny (Tolman i Honzik 1930). W dodatku szczur reagował zaskoczeniem, gdy pokonał tę samą drogę, która dotąd prowadziła do pokarmu, ale tym razem pokarmu tam nie było. W pierwszym wypadku szczur *reprezentuje* drogę poprawnie, aby dojść do tego miejsca, gdzie znajdował się dotychczas pokarm. Wówczas można by rzec, że szczur dobiera właściwe środki do zamierzonego celu². W drugim wypadku szczur zachowuje się tak, że nie osiąga zamierzonego celu: idzie w kierunku miejsca, gdzie wcześniej znajdował się pokarm, a teraz go tam nie ma. Bez pojęcia reprezentacji umysłowej, której przykładem może być błędna reprezentacja szczura,

¹ Donald Davidson w *Zwierzętach racjonalnych* (Davidson 1992) argumentuje za tym, że irracjonalność przekonań czy działań może być przypisana tylko pod warunkiem, że dany podmiot ma przeważająco wiele racjonalnych przekonań.

² Tolman bronił odmiany behawioryzmu, którą nazywał „celowościową” (Tolman 1995).

że w danym miejscu jest pokarm, przypisanie racjonalności zwierzęciu byłoby bardzo trudne. Bez pojęcia reprezentacji utrudnione staje się wyjaśnienie, dlaczego szczur zmierza w stronę poprzedniego położenia pokarmu, a przede wszystkim dlaczego jest zaskoczony³.

Powszechnie posługujemy się pojęciem błędnej reprezentacji, gdy opisujemy i wyjaśniamy nasze zachowania, także te, które przybliżają nas do zamierzonego celu, np. Jaś stoi na przystanku o danej porze, gdyż czeka na autobus, ma on jednak błędne przeświadczenie o rozkładzie jazdy i nie wie, że autobus nie przyjedzie. Pojęcie reprezentacji umysłowej pozwala także przewidywać czyjeś zachowania, a także je zmieniać. I tak przewidujemy zachowanie Jasia, gdy znając jego nawyki, zapytawszy go, gdzie idzie o danej porze, spodziewamy się, że Jaś stanie na przystanku autobusowym, bo zamierza dojechać do szkoły. Wpływamy natomiast pośrednio na zachowania Jasia, powstrzymując je lub powodując ich zmianę, gdy zmieniamy jego reprezentacje umysłowe, np. tłumacząc Jasiowi, że patrzy na nieaktualny rozkład. Warto zauważyć, że wpływając na czyjeś reprezentacje umysłowe, można zmienić jego zachowania: dużo więcej wysiłku wymagałoby zaciągnięcie Jasia przemocą na przystanek (przy życzliwym założeniu, że Jaś nie jest bardzo uparty).

Podsumowując: bez pojęcia błędnej reprezentacji nie można systematycznie wyjaśniać zachowań w kategoriach racjonalności instrumentalnej, gdyż należałoby założyć, iż z niewiadomych powodów racjonalne systemy poznawcze powszechnie popełniają błędy, dobierając niewłaściwe środki do swoich celów. Co więcej, tego rodzaju rzekomej irracjonalności zachowania nie można by usunąć, poprawiając wnioskowania systemu poznawczego, gdyż z założenia system ten miałby wyłącznie poprawne reprezentacje. Jedynym praktycznym sposobem usunięcia tajemniczej irracjonalności byłaby bodaj interwencja neurologiczna. Cóż bowiem innego mielibyśmy zrobić, jeśli Jaś jest z zasady nieomylny? Przy założeniu, że nie istnieją reprezentacje błędne, system poznawczy mający reprezentacje umysłowe musi je mieć wszystkie poprawne (jeśli w ogóle je ma): jeśli system coś reprezentuje, to nieomylnie. Ponieważ założenie o ludzkiej nieomyślności jest bardzo mało wiarygodne, odrzucam propozycję wyrugowania pojęcia błędnej reprezentacji z koncepcji reprezentacji umysłowych. Właśnie do takich niepożądanych konsekwencji

³ Przeciwnicy Tolmana długo starali się opisywać zachowanie szczurów w kategoriach czysto behawioralnych, bez reprezentacji umysłowych. Jak wskazuje jednak William Bechtel (Bechtel 2014), nigdy nie mogli przewidzieć ciekawych nowych zjawisk ani projektować nowych eksperymentów. Mocniejsza hipoteza Tolmana ma bowiem większą moc heurystyczną i stanowi zarzewie płodnego do dzisiaj programu badawczego. W 2014 r. odkrycia Johna O'Keefe'a i małżeństwa Moserów dokonane w ramach tego programu uhonorowano Nagrodą Nobla.

prowadzi m.in. ujęcie Marka Perlmana, który argumentuje na rzecz eliminacji tego pojęcia (Perlman 2000).

Pojęcie błędnej reprezentacji odgrywa rolę nie tylko w opisywaniu, wyjaśnianiu i przewidywaniu zachowań pozornie irracjonalnych, lecz także w odniesieniu do (przynajmniej pewnego rodzaju) uczenia się na błędach. Uczenie tego rodzaju zakłada, że organizm jest w stanie *rozpoznać* swój błąd. Załóżmy, że Staś rozwiązuje zadania z arytmetyki ze zbioru zadań, do których odpowiedzi znajdują się na końcu książki. Staś wie, że gdzieś popełnił błąd w zadaniu i dlatego stara się usunąć przyczynę tego błędu, próbując wykonać działanie w taki sposób, żeby pasowało do odpowiedzi. Uczenie się na błędach jest powszechnie występującym zjawiskiem, także w wypadku bardziej pierwotnych umiejętności motorycznych. Przykładowo, dziecko nieprawidłowo jeździ na rowerze, gdy wykonuje nieprawidłowy plan motoryczny, w wyniku czego upada na ziemię. Nauka na błędach polega na wyuczeniu się prawidłowej sekwencji ruchów, dzięki czemu dziecko przestanie się przewracać.

W pracy przyjmuję założenia naturalizmu. Naturalizm to stanowisko, według którego: 1) istnieje tylko rzeczywistość naturalna, na którą nie wpływają żadne obiekty ponadnaturalne, a w związku z tym 2) wszystkie aspekty rzeczywistości można badać metodami naukowymi. Wyróżnione numerami tezy naturalizmu odpowiadają kolejno jego dwóm aspektom, w związku z czym wyróżnia się jego dwa podrodzaje: naturalizm (1) ontologiczny i (2) metodologiczny (Papineau 2015). W książce przyjmuję perspektywę naturalizmu metodologicznego, który jest stanowiskiem powszechnym we współczesnej filozofii, zwłaszcza we współczesnej filozofii nauki. Naturalizm metodologiczny jako poprawne metody uzasadniania twierdzeń dopuszcza wyłącznie metody współczesnej nauki oraz wszelkie rzetelne i intersubiektywnie sprawdzalne metody, które mogą być stosowane w nauce i których wyniki są przez te metody uzasadnialne⁴. Jako taki naturalizm metodologiczny nie musi ograniczać się do metod już istniejących i dojrzałych nauk; jeśli tylko wykaże się zgodność proponowanej metody badawczej z przyjętym już zespołem uznanych za poprawne praktyk badawczych w nauce, to wyjaśnienie stosujące tę metodę jest zgodne z naturalizmem metodologicznym. Ponieważ filozofia często inspiruje powstawanie dyscyplin badawczych, a także sugeruje nowe metody badawcze, filozofowie często są twórcami

⁴ Stwierdzenie, czym są metody rzetelne i intersubiektywnie sprawdzalne, jest bardzo skomplikowaną sprawą. Naturalizm metodologiczny zakłada, że rzetelność i intersubiektywność metod badawczych podlega sprawdzaniu wewnątrz samej nauki, która przyjmuje nowe metody, posługując się metodami wcześniej przyjętymi. Nauka ciągle rozbudowuje i zmienia repertuar swoich metod, przez co jej stan można porównać do przebudowującej się na morzu łodzi, zgodnie ze znaną metaforą Ottona Neuratha (Neurath 1983).

nowych metod nauki. Na przykład William James w dziedzinie psychologii zaproponował badanie emocji przez badanie stanów ciała, opierając się na swoich argumentach filozoficznych⁵.

Antynaturalizm metodologiczny jest to stanowisko, według którego jakaś dziedzina rozważań, np. humanistyka, jest traktowana odrębnie od nauki, ma inne cele badawcze i inne metody⁶. Przykładem nienaturalistycznych metod w naukach humanistycznych może być wczucie w sensie Wilhelma Diltheya (Dilthey 2004).

Odnieśmy założenie naturalizmu do reprezentacji. Franciszek Brentano uważał, że intencjonalność jest cechą odróżniającą umysły od świata fizycznego (Brentano 1999). Dzięki niemu pojęcie intencjonalności powróciło do współczesnej filozofii. Brentano rozumiał „intencjonalność” jako odniesienie do pewnej treści, skierowanie na pewien obiekt lub immanentną przedmiotowość. W ujęciu tego filozofa intencjonalność jest swoista dla tego, co wyłącznie umysłowe, w przeciwieństwie do tego, co umysłowe nie jest. Innymi słowy, intencjonalność jest przeciwstawiana temu, co należy do świata materialnego.

⁵ James argumentował, że emocje sprowadzają się do stanów cielesnych. Jego argument przyjmuje formę eksperymentu myślowego polegającego na odejmowaniu cielesnych cech z odczuwania emocji. W wyniku tego eksperymentu okazało się, że po odjęciu cielesnych cech odczuwania emocji nic już nie pozostaje: „Jeżeli wyobrazimy sobie jakąś silną emocję, a następnie staramy się wyabstrahować ze swej świadomości tej emocji wszystkie doznania związane z jej symptomami cielesnymi, to stwierdzimy, że nie ma tam nic więcej” (James 2002: 348). Tego rodzaju argumentacja umożliwia psychologom pomiar emocji przez pomiar stanu ciała, co jest przykładem metody badawczej.

⁶ Warto zaznaczyć, że naturalizm współczesny różni się od naturalizmu XIX-wiecznego, w którym pomijało się aspekty kulturowe zjawisk znaczeniowych redukowanych do zjawisk czysto biologicznych lub ekonomicznych. Naturalizm metodologiczny nie wyklucza traktowania racji jako przyczyn zachowania, zwłaszcza w ekonomii klasycznej (Ross 2014), a jako taki pozostaje zgodny z naturalizmem antypozytywistycznym Poznańskiej Szkoły Metodologicznej (Kmita 1971: 175–182). Naturalizm antypozytywistyczny przyjmuje istnienie pojęć idealizujących („typowo-idealnych”) zarówno w naukach humanistycznych, jak i w przyrodoznawstwie, co prowadzi do odrzucenia pozytywistycznego indukcjonizmu w obu dziedzinach nauki. Teza o antyindukcjonizmie i teza o racjonalizującym charakterze interpretacji humanistycznej powoduje, że naturalizm antypozytywistyczny odrzuca nie tylko indukcjonizm, ale również fenomenalizm oraz psychologizm. Jako taki jest radykalnie odmienny od naturalizmu pozytywistycznego. W przeciwieństwie jednak do naturalistów antypozytywistycznych Szkoły Poznańskiej naturaliści metodologiczni dopuszczają badania neuropsychologiczne czy behawioralne nad motywacjami. Ponadto, w przeciwieństwie do antynaturalistów Szkoły Poznańskiej, naturaliści metodologiczni zazwyczaj nie uznają autonomii nauk humanistycznych. Naturalizm metodologiczny nie pociąga wszakże skrajnego redukcjonizmu biologicznego czy (neuro)psychologicznego. Redukcjonizm w odniesieniu do humanistyki głosi, że wszystkie zjawiska kulturowe i znaczeniowe dadzą się sprowadzić do biologii lub (neuro)psychologii, skrajny redukcjonizm zaś głosi, że niektóre terminy oznaczające zjawiska kulturowe bądź znaczeniowe są zastępowalne przez terminy biologiczne lub (neuro)psychologiczne, a inne należy wyeliminować z teorii.

Takie dualistyczne ograniczenie pojęcia intencjonalności do tego, co umysłowe, jest niezgodne z założeniem o naturalizmie przyjmowanym powszechnie w badaniach naukowych nad poznaniem, że intencjonalność nie jest pojęciem fundamentalnym w takim sensie, w jakim fundamentalne są pojęcia podstawowych i uniwersalnych teorii fizycznych. Z perspektywy naturalistycznej niezbędna jest eksplikacja pojęcia intencjonalności w kategoriach innych niż intencjonalność sama, czyli w kategoriach pojęć już stosowanych w nauce lub takich, które można by w tych kategoriach wyeksplikować. Tak samo niezbędna i zasadna jest eksplikacja pojęcia reprezentacji umysłowej w kategoriach niereprezentacyjnych.

Założenie o naturalizmie metodologicznym pociąga:

Metadezyderat I: Adekwatna teoria reprezentacji jest związana z badaniami nauk empirycznych.

Sformułowanie metadezyderatu I dotyczy tego, by teoria reprezentacji była zgodna z wynikami i metodami nauk empirycznych, a więc by założenia samej teorii nie odbiegały od założeń przyjmowanych w naukach empirycznych. A przecież wśród naturalistycznie zorientowanych filozofów są także tacy, którzy nie respektują związku z badaniami nauk empirycznych – to wszyscy tzw. naturaliści nieredukcyjni, a więc ci, którzy dodają nowe byty do ontologii bez teoretycznych racji wyrastających z badań empirycznych. Na przykład Alvin Goldman (Goldman 1993), Galen Strawson (Strawson 2009) czy John Searle (Searle 1983) uważają świadomość fenomenalną za podstawę intencjonalności. Takie stany fenomenalne to nowe byty wprowadzane bez związku z badaniami naukowymi, a jedynie ze względu na założenia filozoficzne.

Teoria reprezentacji powinna móc także wyeksplikować, na czym polega intencjonalność reprezentacji. Tylko wówczas naukowe pojęcie reprezentacji umysłowej będzie zgodne z tym, jak posługujemy się nim w praktyce. Bez przyjęcia tego założenia nawet skrajnie redukcjonistyczne ujęcie intencjonalności jako np. relacji przyczynowej byłoby całkowicie zadowalające. Formułuję zatem następujący metadezyderat:

Metadezyderat II: Adekwatna teoria reprezentacji może wyeksplikować, na czym polega reprezentacyjny charakter pojęć intencjonalnych w praktyce potocznej i naukowej.

Przez „eksplikację” rozumiem procedurę mającą na celu dokonanie analizy pojęciowej (w odróżnieniu od wyjaśniania, którego przedmiotem są pewne fakty – wyjaśniane przez inne fakty, np. prawa przyrody czy

przyczyny). Tak więc na drodze analizy pojęć reprezentacyjnych, popartej przykładami z praktyki naukowej, powinno stać się jasne, na czym polega ich reprezentacyjność. I tak pojęcie świadomości fenomenalnej nie pozwala zadowalająco wyeksplikować pojęcia intencjonalności, gdyż pojęcie świadomości fenomenalnej jest w tych teoriach równie lub wręcz bardziej niejasne niż pojęcie intencjonalności.

Drugim założeniem filozoficznym w książce jest założenie o realizmie w kwestii istnienia reprezentacji umysłowych. Realizm pozostaje niezgodny z antyrealizmem: antyreprezentacjonizmem, eliminacjonizmem i instrumentalizmem. Antyreprezentacjonizm ontyczny (Hutto i Myin 2013b) głosi, że reprezentacje nie istnieją; antyreprezentacjonizm epistemiczny (Chemero 2000) – że nie mają żadnej funkcji wyjaśniającej. Szczególną odmianą antyreprezentacjonizmu jest eliminacjonizm (Churchland 2002), głoszący, że reprezentacje mogą, a nawet powinny zostać wyeliminowane z przyszłej, kompletnej teorii psychologicznej czy neurokognitywnej. Instrumentalizm (Dennett 1971) z kolei głosi, że pojęcie reprezentacji ma co prawda funkcję opisową, predykcyjną i wyjaśniającą, lecz nie musi być stosowane z założeniem, że cokolwiek realnego desygnuje. Wbrew antyreprezentacjonizmowi ontycznemu realizm przyjmuje istnienie reprezentacji; wbrew eliminacjonizmowi i antyreprezentacjonizmowi epistemicznemu – że mają one funkcję opisową i eksplanacyjną w dojrzałych teoriach kognitywnych; wbrew zaś instrumentalizmowi – że pojęcie to dosłownie desygnuje pewne obiekty, mające określone role przyczynowe w systemach poznawczych.

W książce argumentuję, że można pokazać pewne wspólne cechy pojęcia reprezentacji umysłowej występującego zarówno w praktyce potocznej, jak i badaniach naukowych. Czasem przejawia się ono tylko w postaci szczegółowych pojęć intencjonalnych, takich jak „myśl” czy „postrzeżenie”. Zarysowuję cechy charakterystyczne prototypowego pojęcia reprezentacji. Nie chcę przesądzać, jakie rodzaje reprezentacji istnieją, jaka jest ich natura ani czym się od siebie różnią. Nie zakładam, że prototypowe pojęcie reprezentacji umysłowej jest tożsame z filozoficznymi pojęciami przekonania czy pragnienia, lecz zakładam, że podziela przynajmniej niektóre jego cechy. Zakładam też, że jeśli przekonania i pragnienia istnieją, to są reprezentacjami. Wbrew klasycznemu ujęciu reprezentacji umysłowej nie ograniczam się też do reprezentacji takich jak przekonania i pragnienia wyrażalne językowo, jak czyni to np. Donald Davidson.

W dyskusji nad reprezentacjami umysłowymi naturalizacja intencjonalności i związana z nią naturalizacja treści reprezentacji stanowi szczególną trudność dla reprezentacjonistów (Hutto i Myin 2013a). Dlatego w pracy koncentruję się na naturalizacji intencjonalności.

To właśnie zwolennicy reprezentacjonizmu, którzy podejmują się naturalizacji intencjonalności, eksplikując ją w kategoriach innych niż intencjonalność sama, m.in. Ned Block, Fred Dretske, Ruth Garrett Millikan, będą omawiani w książce.

Ze swoich rozważań w pracy wykluczam dwa rodzaje stanowisk. Po pierwsze te, które zakładają realizm co do intencjonalności, lecz antyrealizm co do reprezentacji umysłowych. Przykładem jest koncepcja Daniela Hutta i Erika Myina. W pracy zajmują się realistycznym ujęciem stanów intencjonalnych i reprezentacji umysłowych. Choć samego anty-reprezentacjonizmu Hutta i Myina nie analizuję szczegółowo, odnoszę się do przedstawionego przez nich tzw. trudnego problemu treści (patrz niżej), który stanowi silny argument za odrzuceniem realizmu. Po drugie, wykluczam stanowiska, które wprowadzie przyjmują intencjonalność i realizm w kwestii reprezentacji umysłowych, lecz nie naturalizują intencjonalności. Przykładem jest koncepcja Johna Searle'a (Searle 1983).

Argumentuję na rzecz takiej filozoficznej koncepcji reprezentacji umysłowych, która mogłaby stanowić metateorię dla empirycznych teorii reprezentacji umysłowych. Sama koncepcja filozoficzna nie wystarczy do spełnienia proponowanych dezyderatów, gdyż dodatkowo potrzebne jest po prostu zbudowanie adekwatnej empirycznej teorii czy modelu. Ponieważ przyjmuję założenia naturalizmu metodologicznego, koncepcja ta powinna spełnić te dezyderaty, odpowiednio opisując strukturę zjawisk związanych z reprezentacjami. Dalej w tekście książki będę dla uproszczenia pisać również o koncepcjach filozoficznych reprezentacji, że spełniają bądź nie takie a takie dezyderaty; będzie to wówczas oznaczało, że empiryczna teoria lub model wykorzystujący taką koncepcję spełniłyby moje dezyderaty. Sama koncepcja filozoficzna dosłownie nie może ich spełnić; może dostarczyć odpowiedniego instrumentarium pojęciowego. Nie dostarczam empirycznej teorii reprezentacji umysłowych, gdyż to nie jest rolą filozofa.

Celem książki jest uzasadnienie i uporządkowanie wymogów, jakim powinna sprostać satysfakcjonująca koncepcja filozoficzna reprezentacji umysłowych, która jest zgodna z praktyką potoczną i z badaniami naukowymi⁷. Analizuję wybrane współczesne koncepcje reprezentacji

⁷ Można sobie wyobrazić co najmniej dwa sposoby realizacji tego celu. Pierwszy z nich sprowadzałby się do stworzenia teorii metateorii, a więc sformułowania *a priori* wymogów wobec filozoficznej koncepcji reprezentacji umysłowych, a następnie ich uporządkowania i uzasadnienia. Nie jest to moim zamiarem. Inny sposób realizacji tego celu jest *post hoc*: polega on na sformułowaniu dezyderatów, ich uporządkowaniu i uzasadnieniu dopiero w wyniku systematycznego przeglądu współczesnych filozoficznych koncepcji reprezentacji umysłowych przy założeniu naturalizmu i realizmu. Przegląd koncepcji dostarcza zarazem argumentacji na rzecz takiej koncepcji, która spełniałaby

umysłowych pod kątem wprowadzonych metadezyderatów i odpowiadających im dezyderatów oraz omawiam wymogi wobec filozoficznych koncepcji poszczególnych rodzajów reprezentacji umysłowych. W efekcie opracowuję własną naturalistyczną koncepcję reprezentacji umysłowych spełniającą, w moim przekonaniu, wszystkie dezyderaty i zapewniającą właściwe ujęcie błędu reprezentacyjnego.

Jednym z głównych problemów koncepcji reprezentacji jest ujęcie błędu reprezentacyjnego. Nie wszystkie proponowane współcześnie koncepcje podejmują ten problem, a nawet jeśli tak się dzieje, to nie wypracowują ujęcia błędu dla systemu. Szczególnym efektem tej książki jest wypracowanie autorskiej konceptualizacji błędu reprezentacyjnego rozpoznawalnego przez system poznawczy w ramach koncepcji teleosemantycznej.

Książka ma charakter metodologiczno-teoretyczny. Posługuję się metodą analizy pojęciowej. Przez analizę pojęciową zwykle rozumie się dostarczenie warunków koniecznych i wystarczających stosowania danego pojęcia i tym samym zbudowanie definicji normalnej (równościowej lub równoważnościowej) danego pojęcia. W książce podaję nie tyle definicję normalną pojęcia reprezentacji umysłowej, ile jej definicję uwikłaną⁸. Sporządzam listę wybranych wymogów wobec teorii reprezentacji umysłowych, które nazywam „dezyderatami wobec adekwatnej teorii reprezentacji umysłowej”. Wybrane dezyderaty uzasadniam i porządkuję.

Ze względu na rolę przykładów analiza pojęciowa dokonywana w tej książce przypomina raczej analizę trójdzielną (*3-analysis*) w sensie Paula Thagarda (Thagard 2015), który rozróżnia trzy elementy takiej analizy: 1) dostarczenie prototypowych przykładów, 2) wskazanie typowych cech, 3) opis charakterystycznych wyjaśnień z użyciem danego pojęcia. Z jednej strony więc przyglądam się przykładom z praktyki potocznej i z badań naukowych, a z drugiej – wybieram tylko niektóre użycia pojęcia reprezentacji umysłowej jako takie, które *prima facie* spełniają wyróżnione przeze mnie dezyderaty wobec pojęcia reprezentacji. Paradigmatyczne przykłady użycia pojęcia reprezentacji umysłowej traktuję poważnie właśnie dlatego, że przyjmuję stanowisko naturalistyczne, zgodnie z którym filozofia nie ma uprzywilejowanego punktu widzenia, pozwalającego jej rozstrzygać wszelkie spory teoretyczne, lecz musi odwoływać się do tego,

wyszczególnione dezyderaty. W tym sensie stworzenie takiej koncepcji wiąże się z uzasadnieniem i uporządkowaniem wymogów, jakim powinna sprostać satysfakcjonująca koncepcja filozoficzna reprezentacji umysłowych.

⁸ Jak argumentuje np. Jerry Fodor (Fodor 1980), nie jesteśmy w stanie podawać jawnych definicji, a więc definicji konotacyjnych, jakichkolwiek elementarnych pojęć empirycznych. Zdarzają się oczywiście definicje konotacyjne pojęć matematycznych (np. „Kwadrat to prostokąt o wszystkich bokach równych”).

co jest obecnie uznawane za najlepiej uzasadnione. Taką właśnie cechą mają przywoływane przeze mnie paradygmatyczne przykłady. Filozofia nie tylko ma formułować normy rządzące stosowaniem pojęcia reprezentacji, lecz także poprawnie opisywać jego użycie.

Rolą analizy przykładów jest dalsze uzasadnienie, dlaczego pojęcie reprezentacji umysłowej zawężam w taki, a nie inny sposób. Analizując przykłady, wskazuję na swoistą rolę, jaką reprezentacje umysłowe mają – podobnie jak pojęcie dobra moralnego ma swoje nieodzowne zastosowanie w etyce do opisu i wyjaśniania klasy wartości pozytywnych, tak pojęcie reprezentacji umysłowej ma swoje nieodzowne zastosowanie w filozofii i kognitywistyce, gdyż pozwala wyróżnić pewną klasę procesów poznawczych. Innymi słowy, pokazuję, w jakim sensie pojęcie reprezentacji nie jest trywialne, co pozwala spełnić wymóg Ramseya (Ramsey 2007) opisu zadania wobec teorii eksplikującej pojęcie reprezentacji (*job description challenge*). Na wybranych przykładach z etologii i psychologii pokazuję, że można mówić o pewnym jednolitym pojęciu reprezentacji umysłowej u człowieka i u zwierząt, które spełnia wyróżnione przeze mnie dezyderaty.

Dezyderaty i metadezyderaty, które proponuję, powinny być spełniane przez empirycznie sprawdzalne teorie lub modele⁹ wykorzystujące pojęcie reprezentacji umysłowej. Ponieważ wszystkie takie modele i teorie opierają się na pewnym aparacie teoretycznym, spełnienie dezyderatów zależeć będzie także od tego, czy ten aparat jest z nimi zgodny pod względem metodologicznym. Na przykład behawioryzm Skinnera (Skinner 2013) żadnych moich dezyderatów nie spełni: zachowanie wyjaśnia się w nim tylko w kategoriach bodźców i reakcji, bez postulowania jakiegokolwiek struktury systemu poznawczego, a więc i bez reprezentacji.

Metadezyderaty prowadzą do coraz bardziej szczegółowych dezyderatów. Podam teraz uzasadnienie proponowanej ich listy. Pełniejszego uzasadnienia dostarczam przy omawianiu poszczególnych koncepcji filozoficznych i przy analizie przykładów z etologii i psychologii.

Argumentuję za tym, że pojęcie reprezentacji umysłowej powinno odgrywać różne role, takie jak opisywanie, wyjaśnianie i przewidywanie zachowań. Uzasadniam to tym, że tylko takie pojęcie reprezentacji umysłowej jest zgodne z praktyką potoczną i badaniami naukowymi. Dodatkowo zaznaczam, że przez zmianę stanów intencjonalnych można zmieniać zachowania systemów poznawczych, czyli interweniować w ich zachowanie. Formułuję w związku z tym dezyderat eksplanacyjny:

⁹ Pojęć „teoria” i „model” używam tak, jak są one stosowane w kognitywistyce. Teoria opisuje klasę zjawisk (np. wyjaśnianych lub przewidywanych). Model opisuje zaś konkretne zjawiska lub grupę konkretnych zjawisk, zwykle korzystając z ogólniejszej teorii.

I.1. Adekwatna teoria reprezentacji uwzględnia ich rolę w opisie, wyjaśnianiu, przewidywaniu procesów poznawczych i zachowań oraz w interweniowaniu w procesy poznawcze i zachowania.

Teoria reprezentacji powinna ujmować najszerszy zakres procesów poznawczych i zachowań, w taki sposób jednak, żeby uwzględnić swoistość zjawisk reprezentacyjnych i jednocześnie nie doprowadzić do trywializacji pojęcia reprezentacji umysłowej. Nie powinno się *a priori* zawężać pojęcia reprezentacji tylko do zachowań ludzkich czy tylko do zachowań zwierzęcych ani tylko do zachowań ujmowanych na poziomie osobowym czy tylko do procesów zachodzących na poziomie subosobowym. Myśl tę wyrażają dezyderaty o uniwersalności (I.2) i dezyderat o wielopoziomowości (I.3):

I.2. Adekwatna teoria reprezentacji dotyczy zarazem zjawisk językowych i niejęzykowych (np. percepcyjnych) tak ludzi, jak i zwierząt.

I.3. Adekwatna teoria reprezentacji dotyczy zjawisk ujmowanych zarówno na poziomie osobowym, jak i subosobowym.

Dezyderaty te uprawomocniają fakt, że ludzie także są zwierzętami, dlatego adekwatna teoria reprezentacji powinna uwzględnić nie tylko ludzkie zachowania i procesy, a w szczególności nie tylko procesy poznawcze związane z językiem.

Pojęcie reprezentacji umysłowej wiąże się z racjonalnością. W świecie biologicznym spotyka się przejawy racjonalności instrumentalnej. Ujmuję to jako dezyderat o racjonalności:

II.1. Adekwatna teoria reprezentacji jest zgodna z założeniami dotyczącymi racjonalności systemu poznawczego.

W książce nie dostarczam konkretnej teorii racjonalności, nie eksplikuję też dokładnie związku racjonalności z reprezentowaniem. Argumentuję jednak za tym, że teoria reprezentacji umysłowych, która opisuje zachowania ludzi i zwierząt, nie powinna prowadzić do konsekwencji, że system poznawczy popełniający błędy jest całkowicie irracjonalny.

Pojęcie reprezentacji umysłowej wiąże się z odzwierciedleniem rzeczywistości. Ma to uzasadnienie biologiczne, gdyż życie wyłącznie w świecie fikcji lub wyobrażeń nie pozwoliłoby organizmom przetrwać. Jako takie pojęcie reprezentacji umysłowej nie powinno dotyczyć wyłącznie reprezentacji pustych ani fikcyjnych. Tę ideę wyraża dezyderat

o naturalizacji intencjonalności, który rozбивa się na dezyderaty o determinacji odniesienia i treści oraz dezyderat o nośniku treści:

II.2. Adekwatna teoria reprezentacji naturalistycznie eksplikuje pojęcie intencjonalności, czyli

II.2.1. wskazuje determinanty odniesienia i treści;

II.2.2. pokazuje, dlaczego reprezentacja może być nośnikiem treści.

W klasycznych koncepcjach filozoficznych, jak np. Johna Stuarta Milla czy Gottloba Fregego, reprezentacje mają nie tylko odniesienie, ale i treść. Pojęcie reprezentacji umysłowej, które uwzględnia wyłącznie odniesienie, może desygnować wyłącznie proste przedmioty lub własności. Fakt, że reprezentacje mogą być nośnikami treści, jest przydatną dla organizmów cechą reprezentacji. Treścią są warunki spełniania, które w wypadku bardziej złożonych deklaratywnych reprezentacji językowych są tożsame z warunkami prawdziwości. Gdyby reprezentacje nie mogły mieć warunków spełniania, nie mogłyby być pod tym względem oceniane, a więc organizmy nie mogłyby wnioskować o wartości poznawczej reprezentowanych przedmiotów lub cech. Koncepcja reprezentacji umysłowych, która nie spełnia dezyderatu nośnika treści, a więc nie naturalizuje pojęcia treści, nie rozwiązuje tzw. trudnego problemu treści (Hutto i Myin 2013a). Zdaniem Hutto i Myina problem jest trudny, gdyż w naturalistycznych koncepcjach intencjonalności problematyczne jest wyeksplikowanie pojęcia warunków spełniania.

Produktywność jest własnością struktury systemu reprezentacji, umożliwiającą wytwarzanie potencjalnie nieskończenie wielu reprezentacji. Zdaniem Fodora jedynym wiarygodnym wyjaśnieniem produktywności systemu reprezentacji jest jego kompozycyjność i systematyczność. Jeśli system reprezentacji cechuje się kompozycyjnością, to reprezentacje łączą się ze sobą w bardziej złożone reprezentacje lub rozkładają się na prostsze. W przeciwieństwie do kompozycyjności pojęcie systematyczności definiowane jest w sposób intuicyjny (Aizawa 2003). I tak system reprezentacji jest systematyczny, jeśli można w nim wytworzyć zarówno reprezentację „Jaś kocha Marysię”, jak i „Marysia kocha Jasia”. Systematyczność jest to więc taka własność systemu reprezentacji, która umożliwia powstawanie reprezentacji w taki sposób, że jeśli może powstać reprezentacja o pewnej strukturze, to powstać mogłaby też reprezentacja o innej, systematycznie zmienionej strukturze. Niestety w literaturze przedmiotu nie określono ściśle katalogu tych zmian, które pozwalałyby sprecyzować tę nieco nieostrą definicję. Podsumowując, warunkiem

tę, żeby system reprezentacji był produktywny, jest systematyczność i kompozycyjność.

Adekwatna teoria reprezentacji powinna dostarczyć właściwej koncepcji treści, co wyraża dezyderat o takiej nazwie: dezyderat o właściwej koncepcji treści (II.3) i podpadające pod niego dezyderaty o produktywności (II.3.1), a także dwa dezyderaty o uwzględniania błędu (kolejno: II.3.2 o możliwości błędu i II.3.3 o błędzie wykrywalnym przez system):

II.3. Adekwatna teoria reprezentacji dostarcza właściwej koncepcji treści (w tym opisuje związek intensji i ekstensji jako nieprzypadkowy).

II.3.1 Adekwatna teoria treści reprezentacji uwzględnia produktywność części reprezentacji.

II.3.2. Adekwatna teoria treści reprezentacji uwzględnia możliwość błędnej reprezentacji.

II.3.3. Adekwatna teoria treści reprezentacji uwzględnia możliwość rozpoznawania adekwatności reprezentacji przez system poznawczy.

Dezyderat o produktywności uzasadniam tym, że adekwatna teoria reprezentacji umysłowej powinna ujmować strukturalne własności reprezentacji, dzięki którym część z nich jest produktywna. Dotyczy to w szczególności reprezentacji pojęciowych i językowych (Aizawa 2003). Dezyderat o możliwości błędu uzasadniam tym, że uwzględnienie możliwości błędnej reprezentacji pozwala odróżnić reprezentację błędną od braku reprezentacji. Adekwatna reprezentacja umysłowa powinna móc być wartością dla systemu poznawczego. Dzieje się tak wtedy, gdy ma ona charakter poznawczy dla tego systemu, czyli organizm jest w stanie odróżniać (przynajmniej niektóre) reprezentacje błędne od poprawnych. Dlatego przyjmuję dezyderat o błędzie wykrywalnym przez system – według tego dezyderatu adekwatna teoria treści powinna uwzględnić możliwość rozpoznawania adekwatności reprezentacji przez system poznawczy. Oba wymogi są kluczowe, gdyż wskazują na to, że dzięki reprezentacjom systemy poznawcze mogą popełniać błędy; dezyderat o wykrywalności błędu przez system jest szczególnie istotny, gdyż mówi o swoście reprezentacyjnej roli reprezentacji umysłowych.

W części pierwszej wychodzę od obserwacji, że pojęcia intencjonalne, m.in. przekonania i pragnienia, ale nie tylko one, są stosowane powszechnie w praktyce potocznej. Przedstawiam w związku z tym koncepcję Dennetta nastawienia intencjonalnego, gdyż broni on sze-

rokiego stosowania takich pojęć. Zwracam także uwagę na zgodność wyjaśnień w kategoriach reprezentacji umysłowych z zasadami racjonalności i życzliwości. Koncepcję nastawienia intencjonalnego traktuję więc jako poprawną analizę filozoficzną praktyki potocznej.

W rozdziale tym pokazuję także, w jaki sposób pojęcia intencjonalne wchodzą w związek z racjonalnością. Pozostając w zgodzie z Dennettowską koncepcją nastawień, polemizuję z argumentem Davidsona przeciw racjonalności zwierząt. Wynikiem rozdziału 1. jest obrona pojęcia reprezentacji umysłowej jako pojęcia intencjonalnego, pozostającego w zgodzie z praktyką potoczną i badaniami naukowymi. Argumentuję w tym rozdziale przede wszystkim na rzecz dezyderatów: ekspanacyjnego (I.1), uniwersalności (I.2) i wielopoziomowości (I.3), a także o racjonalności (II.1).

W części drugiej koncentruję się na najprostszych koncepcjach reprezentacji umysłowych, które dotyczą nie tylko poziomu całych organizmów, lecz także subosobowych procesów poznawczych. Przyglądam się wyróżnionym stanowiskom funkcjonalistycznym. W rozdziale tym omawiam koncepcję czysto internalistyczną. Pokazuję, że internalizm ma pewne zalety, gdyż podkreśla wewnętrzny wymiar systemu poznawczego, koncentrując się na intensjonalnej charakterystyce reprezentacji. Przyjmuję pewne racje ekspanacyjne koncepcji treści wąskiej (za Fodorem i Segalem), pokazując, dlaczego pojęcie treści jest istotne w teoriach reprezentacji umysłowej, lecz twierdząc, że pojęcie treści wąskiej nie nadaje się do ekspanacji pojęć stosowanych w etologii i psychologii poznawczej. W dalszej części książki ograniczam się więc do eksternalistycznych koncepcji reprezentacji umysłowych.

W rozdziałach 3. i 4. analizuję koncepcje przyczynowo-informacyjne, koncentrujące się wyłącznie na odniesieniu, a pomijające treść. Są to: semantyka przyczynowa w stylu Wisconsin i semantyka warunków prawdziwości Davidsona. Pokazuję, że koncepcje te ograniczają się do odniesienia lub warunków prawdziwości, z czym wiążą się pojęciowe kłopoty.

Podsumowując część drugą, stwierdzam, że żadna z uproszczonych teorii reprezentacji umysłowej – ani skrajny internalizm, ani skrajny eksternalizm – nie zdaje sprawy z potocznego użycia pojęcia reprezentacji umysłowych, które nakazuje uwzględniać zarówno rzeczywistość poza umysłem, jak i wewnętrzną perspektywę systemu poznawczego. Nie zdaje również sprawy ze związku reprezentacji umysłowych z badaniami naukowymi. I tak brak treści, odniesienia czy weredyczności nie pozwala stosować powszechnych w etologii i psychologii wyjaśnień z użyciem map poznawczych oraz pojęć. Dlatego – zgodnie z teoriami eksternalistycznymi – stwierdzam, że teorie reprezentacji umysłowych powinny uwzględnić odniesienie i weredyczność, aby móc spełnić dezyderat o naturalizacji intencjonalności.

Pokazuję także, że czysto eksternalistyczna teoria reprezentacji umysłowej ma pojęciowe trudności z determinacją odniesienia, których nie może przezwyciężyć bez pojęcia treści. Argumentuję na rzecz konieczności uwzględnienia zarówno czynnika treściowego, jak i odniesieniowego (dzięki czemu teoria ta będzie mogła spełnić dezyderat o możliwości błędu). Dwuczynnikowa koncepcja Blocka służy mi jako przykład nieprawidłowego związku ekstensji i intensji, świadczący o tym, że ów związek nie może polegać tylko na koniunkcji odrębnych teorii treści i teorii odniesienia.

Dopiero eksternalistyczne koncepcje teleosemantyczne (z części trzeciej) odpowiednio ujmują związek treści i odniesienia, gdyż dostarczają pojęcia treści szerokiej. Teleosemantyka uzasadnia związek treści i odniesienia na gruncie koncepcji funkcji biologicznej systemów poznawczych. Systemy biologiczne wykorzystują reprezentacje umysłowe do lepszej orientacji w świecie, niezbędnej dla przetrwania i adaptacji do środowiska. Przede wszystkim jednak teleosemantyczna koncepcja funkcji w taki sposób pozwala związać treść z odniesieniem, że umożliwia przypisywanie treści reprezentacjom, które są błędne lub poprawne, co jest najważniejszym elementem adekwatnej teorii reprezentacji umysłowych. Tym samym teleosemantyka spełnia pierwszy z kluczowych dezyderatów o możliwości błędu.

Warto podkreślić, że dopuszczam teoretyczną możliwość eksplikacji błędu reprezentacyjnego w kategoriach innych niż funkcjonalne. Obecnie jednak bodaj tylko funkcjonalne teorie teleosemantyczne dostarczają zadowalającej eksplikacji tego błędu (asymetryczna koncepcja przyczynowa Fodora (Fodor 1992; 2011), omawiana w części drugiej, pozostaje zbyt ogólnikowa i jego kryterium poprawności nie jest swoiste dla teorii reprezentacji umysłowej). Jest tak, gdyż w teleosemantyce pojęcia takie jak „reprezentowanie”, „błędne reprezentowanie” i „niereprezentowanie” są eksplikowane odpowiednio w kategoriach funkcji, dysfunkcji i braku funkcji.

Zarazem teleosemantyczne koncepcje nie są w stanie poradzić sobie z eksplikacją idei błędu rozpoznawalnego przez system (*błąd dla systemu*), dlatego nie spełniają drugiego kluczowego dezyderatu o wykrywalności błędu przez system. Wyjątkiem jest teoria Marka Bickharda, która jednak ogranicza się do reprezentacji umysłowych rozumianych antycypacyjnie i zawiera mocne metafizyczne założenia procesualizmu. W ostatnim rozdziale książki proponuję autorską koherencyjną koncepcję błędu dla systemu, która nie wymaga antycypacji. Zaproponowana przeze mnie koncepcja bazuje na Millikanowskim modelu reprezentacji i w nowy sposób eksplikuje ideę błędu rozpoznawalnego przez system.

W praktyce naukowej pojęcie reprezentacji umysłowej jest stosowane nie tylko instrumentalnie, zwłaszcza wtedy, gdy teoretycy starają

się określić, dzięki czemu reprezentacje mają taką, a nie inną treść. Odwołam się do trzech klasycznych przykładów opisujących zachowania i uznawanych przez naukowców za niekontrowersyjne przykłady traktowanych realistycznie reprezentacji umysłowych. Pierwszy pochodzi z dziedziny etologii i jest reprezentacją strukturalną (mapa poznawcza), dwa zaś ostatnie – z psychologii poznawczej (klasyczna teoria pojęć, spór o wyobrażenia umysłowe), spośród których jeden jest reprezentacją strukturalną (wyobrażenia umysłowe), a drugi symboliczną (klipiec). Wybrane przykłady nie tylko dostarczają dalszego uzasadnienia dla dezyderatów, ale również pozwalają sprawdzić, czy koncepcje treści są przydatne. Analizując użycia pojęć reprezentacji umysłowych na wybranych przykładach, będę odwoływać się do różnych eksperymentów z wykorzystaniem tych pojęć; wyjątkiem jest pojęcie klipca, ponieważ wiąże się ono nierozzerwalnie z prekursorskimi badaniami Andrzeja Lewickiego nad uczeniem pojęć.

Reprezentacje strukturalne charakteryzują się homomorfizmem struktury nośnika reprezentacji i struktury desygnatu – składnia nośnika informuje o cechach desygnatu reprezentacji (Cummins i Roth 2014; Ramsey 2007; Sloman 1978). Przykładowo, mapa Warszawy pokazująca przystanek autobusowy informuje (a przynajmniej powinna informować) o tym, gdzie przystanek ten znajduje się względem innych obiektów przedstawionych na tej mapie.

Reprezentacje symboliczne charakteryzują się pewną dowolnością relacji desygnatu reprezentacji i jej nośnika, wobec czego struktura nośnika nie informuje o desygnacie. Przykładowo, reprezentacja symboliczna krowy może mieć postać słowa w języku polskim, np. „krowa”, lub angielskim „cow”, ale też może być zapisana za pomocą systemu dziesiętnego. Bez względu jednak na format reprezentacji struktura jej nośnika nie informuje o strukturze desygnatu, w tym wypadku o krowie (krowa nie składa się z liter ani z elementów, które odpowiadałyby literom w jakimkolwiek języku naturalnym).

Mapy poznawcze

Dziś status map poznawczych jako reprezentacji strukturalnych nie budzi takich sporów jak do niedawna. John O’Keefe odkrył w mózgach ssaków komórki miejsca (w latach sześćdziesiątych XX w.), a Edvard i May-Britt Moserowie 30 lat później odkryli komórki siatki. Te pierwsze znajdują się w hipokampie, a te drugie w korze śródwęchowej. Współpraca tych dwóch rodzajów komórek tworzy system orientacji, który można przyrównać do mapy poznawczej. Za wspólne odkrycie O’Keefe i małżeństwo Moserów dostali w 2014 r. Nagrodę Nobla. Odkryte przez noblistów komórki miejsca i siatki stanowią mapę poznawczą w dosłownym

sensie. Jest ona dynamiczną (zmieniającą się w czasie) strukturą odwzorowującą geometrię otoczenia (dokładniej: strukturą czasową pobudzeń neuronów połączonych geometrycznie, więc aktywność neuronów odpowiada strukturze przestrzennej otoczenia)¹⁰. Wcześniej istnienie mapy poznawczej postulował Edward Tolman (Tolman 1948). Hipoteza mapy poznawczej ma też rolę heurystyczną w badaniach neuropsychologicznych¹¹.

Do niedawna status mapy poznawczej był jednak punktem zapalnym między zwolennikami behawioryzmu i reprezentacjonizmu. Idea mapy poznawczej zrodziła się w polemice z wulgarnym behawioryzmem opartym na schemacie bodziec–reakcja. Tolman wykazał, że zwierzę (szczur) może uczyć się bez obecności wzmocnienia¹². Dzięki badaniom Johna O'Keefego i Lynna Nadel okazało się, że mapa poznawcza ma podłoże neurologiczne w mózgu szczura. Jako taka może być zaobserwowana i można ją zlokalizować w hipokampie (O'Keefe i Nadel 1978)¹³. Istnieją też inne badania, dowodzące istnienia pamięci retrospektywnej i prospektywnej nie tylko u gryzoni, ale również u wron (Trojan 2013). Część tych badań dotyczyła zachowania szczura w wodnym labiryncie, w którym szczur miał za zadanie dotrzeć do pokarmu w ciemności, bez wykorzystania bodźców innego rodzaju niż dotyczące własnego ruchu: zmysłu równowagi i kopii poleceń motorycznych wysyłanych do własnych mięśni (Cheung i in. 2012). Pozostawiony w ciemności szczur doskonale radzi sobie ze znalezieniem pokarmu, co dowodzi istnienia u zwierząt

¹⁰ Termin „mapa poznawcza” jest wieloznaczny (Rescorla 2009). „Mapa poznawcza” występuje w różnych sensach: banalnym (mapa poznawcza jest tożsama z jakimkolwiek mechanizmem umysłowym lub neuronalnym, który pozwala zwierzęciu orientować się w otoczeniu); luźnym (mapa poznawcza jest tożsama z reprezentacją umysłową, która odpowiada geometrycznym własnościom otoczenia) oraz dosłownym (mapa poznawcza to reprezentacja umysłowa, która ma te same bazowe własności i mechanizmy reprezentacyjne co zwykła mapa). Psychologowie nie spierają się co do banalnego sensu mapy poznawczej, gdyż jest tautologią, że mają ją wszystkie zwierzęta, które są zdolne do orientacji przestrzennej. Przedmiotem prawdziwych sporów są dwa pozostałe sensy – sens luźny, w którym mapa poznawcza to coś, co mówi o własnościach geometrycznych, oraz sens dosłowny, w którym mapa poznawcza ma własności geometryczne. Luźny sens jest neutralny względem własności nośnika, a więc mapa może przyjmować dowolny charakter: językowy, obrazowy czy strukturalny; sens dosłowny natomiast nakazuje rozumieć mapę poznawczą jako reprezentację strukturalną (Rescorla 2009).

¹¹ Hipoteza mapy poznawczej po dziś dzień prowadzi do wielu ciekawych odkryć (Bechtel 2014).

¹² Tolman wykonał eksperyment, w którym szczury swobodnie eksplorowały labirynt, ucząc się układu jego ramion, mimo że nie znajdowały w nim żadnej nagrody ani też kary.

¹³ Ponadto istnieją dowody empiryczne dotyczące tworzenia się śladów pamięci w mózgu szczura, w czym główną rolę odgrywają znajdujące się w hipokampie komórki miejsca. Badania te wskazują na istotowy związek mapy poznawczej z pamięcią epizodyczną (Eichenbaum i Cohen 2014; Eichenbaum 2004).

pamięci epizodycznej¹⁴. Innymi słowy, przetwarzanie pochodzących z otoczenia danych jest znacznie bardziej skomplikowane i nie daje się przekonująco opisać wyłącznie w kategoriach bodźca, bez odwołania do złożonych procesów pamięci, zwłaszcza jeśli istnieje prostsze alternatywne wyjaśnienie w kategoriach map poznawczych. Co więcej, spór o mapy poznawcze przybiera niekiedy postać fałszywej alternatywy: behawioryści twierdzą, że istotna jest rola bodźca, sądząc, że zwolennicy map umysłowych bagatelizują jego rolę (Benhamou 1996). W rzeczywistości bodźce wzrokowe mają olbrzymie znaczenie przy tworzeniu się i kształtowaniu map poznawczych, czemu zwolennicy map umysłowych nie zaprzeczają.

Podsumowując: argumenty na rzecz istnienia map poznawczych oraz ich roli w opisie, wyjaśnianiu i przewidywaniu zachowań skłaniają do uznania hipotezy map poznawczych za konkurencyjną wobec wyjaśnienia behawiorystów, którzy ograniczają wyjaśnienie orientacji zwierząt w otoczeniu do uczenia się na podstawie bodźców w środowisku.

Współczesne badania nad mapami poznawczymi u szczurów sugerują hipotezę, że bodźce zmysłowe są wykorzystywane do korekcji błędu bieżącej lub przyszłej lokalizacji (Ferbinteanu i Shapiro 2003). Istnieją również dowody na to, że istnieje sygnał negatywności EEG (ERN)¹⁵ u ssaków, który jest objawem tego, że badany orientuje się, iż popełnił błąd (Fiorillo, Tobler i Schultz 2003; Valerio i Taube 2012). Innym dowodem na to, że ssaki orientują się, że się pomyliły, jest jednoczesne wydzielanie się dopaminy w ich mózgu.

Świadczyć by to mogło o tym, że szczury mogą nie tylko błędnie reprezentować swoje położenie w przestrzeni, ale również orientować się, że mają błędną reprezentację tego położenia, aby ją skorygować.

Wyobrażenia umysłowe

Za klasyczny przykład reprezentacji strukturalnej uznaje się wyobrażenia umysłowe (*mental imageries*). *Mental imagery* to współcześnie określenie techniczne odnoszące się do specjalnej klasy procesów umysłowych, polegających na odpowiednim przekształcaniu treści aktualnie realizowanych aktów percepcyjnych. Może to być np. obrót w umyśle figury, która jest właśnie spostrzegana wzrokowo. Wyobrażenia umysłowe są przedmiotem sporu, który wyraża się w pytaniu o to, czy pewne procesy

¹⁴ Na podstawie badań wysnuto wniosek, że reprezentacja w postaci mapy poznawczej służy wyjaśnieniu zjawiska zapamiętywania punktów orientacyjnych w przestrzeni (pojedyncze reprezentacje) oraz struktury geometrycznej otoczenia i zjawiska orientacji w otoczeniu.

¹⁵ Jest to komponent potencjału wywołanego, czyli komponent potencjału elektrycznego rejestrowanego z powierzchni głowy po zadziałaniu odpowiedniego bodźca.

poznawcze opierają się na reprezentacjach strukturalnych, odpowiadających wyobrażeniom, czy też na reprezentacjach symbolicznych i wiedzy ukrytej (składającej się ze znajomości gramatyki języka, intuicyjnej znajomości fizyki świata i podstawowych reguł społecznych) (Thomas 1999).

W słynnym eksperymencie Rogera Shepharda i Jacqueline Metzler (Shepard i Metzler 1971) badanym prezentowano na ekranie komputera rzuty dwóch figur trójwymiarowych na płaszczyznę, a następnie proszono ich o odpowiedź na pytanie, czy figury te są identyczne i jedynie obrócone względem siebie, czy też są różne i jedna jest lustrzanym odbiciem drugiej. Shephard i Metzler odkryli, że czas reakcji potrzebny na podjęcie decyzji jest liniowo proporcjonalny do kąta obrotu z pozycji pierwotnej.

Spór o wyobrażenia dotyczy statusu wyobrażeń umysłowych. I tak w interpretacji zwolenników stanowiska obrazowego, Shepharda i Metzlera, a także Kosslyna (Shepard i Metzler 1980; Kosslyn 1981) i Pinkera (Pinker 1980), badani dokonują w swojej wyobraźni tych samych przekształceń, jakich dokonywaliby na kartce papieru lub przy wykorzystaniu prawdziwych obiektów w przestrzeni. Zwolennicy stanowiska obrazowego postulują więc istnienie wyobrażeń będących reprezentacjami strukturalnymi w wyjaśnieniach dotyczących zdolności poznawczych związanych z obracaniem figur geometrycznych w myślach oraz oceniania odległości w terenie. Przeciwnicy stanowiska obrazowego, tacy jak Zenon Pylyshyn, zaprzeczają, jakoby stwierdzona przez zwolenników stanowiska obrazowego proporcjonalność czasu wykonania do odległości i kąta obrotu tej figury przesądzała, że to jest reprezentacja strukturalna (Shepard i Metzler 1971). Pylyshyn przedstawił alternatywne wyjaśnienie, postulujące reprezentacje symboliczne, które mają naturę językową i powiązane są z wiedzą ukrytą: jego zdaniem reprezentacje obrazowe powstają w wyniku przetwarzania reprezentacji symbolicznych i są wtórne wobec tych ostatnich (Nęcka, Orzechowski i Szymura, 2006).

Spór o wyobrażenia wpisuje się ponadto w bardziej ogólną dyskusję nad autonomią psychologii. Pylyshyn bronił autonomii psychologii, twierdząc, że teoria wyobrażeń umysłowych nie może być potwierdzona lub obalona przez jakiegokolwiek badania mózgu. Kosslyn natomiast odwoływał się do procesów poznawczych także o charakterze mózgowym.

Spór o wyobrażenia formułowano w postaci fałszywej alternatywy: reprezentacje strukturalne przeciwstawiano reprezentacjom symbolicznym, wymagającym postulowania wiedzy ukrytej. Powstaje wrażenie, że oba stanowiska w sporze są wzajemną negacją, podczas gdy zwolennicy reprezentacji strukturalnych mogą bez popadania w sprzeczność korzystać z wiedzy ukrytej, wnioskowań i procesów obliczeniowych w wyjaśnieniu powstawania obrazów (na co zwrócił uwagę Kosslyn (Kosslyn 1994)).

Klipiec

Trzeci analizowany przeze mnie przykład należy do reprezentacji symbolicznych i wywodzi się z klasycznej teorii uczenia się pojęć. Andrzej Lewicki prowadził badania nad nabywaniem sztucznych pojęć, w tym pojęcia przedmiotu zwanego klipcem. Badanie zostało przeprowadzone w następujący sposób:

W eksperymentach Lewickiego [...] materiałem [eksperymentalnym – K.B.] były różnokolorowe karty o różnych kształtach, zawierające w środku małą figurę, której barwa i kształt również ulegały zmianie. Badacz pokazywał kartę wzorcową, mówił, że jest to „klipiec” [...]. Następnie pokazywał kolejną kartę i pytał, czy zdaniem badanego jest to „klipiec”. Gdy odpowiedź była błędna, badacz podawał odpowiedź poprawną i przechodził do następnej karty. Po pokazaniu pierwszej serii kart badacz pokazywał każdą kartę ponownie, prosząc za każdym razem o stwierdzenie, czy jest to egzemplarz pojęcia. Pierwszy zestaw kart był pokazywany tak długo, dopóki badany nie podał wszystkich odpowiedzi poprawnych. Wtedy badacz przystępował do ekspozycji kart kontrolnych, to jest takich, których badany nigdy wcześniej nie widział. Wśród tych kart były zarówno egzemplarze pojęcia, jak i karty, które egzemplarzami nie były. Poprawne odpowiedzi świadczyły o tym, że badani potrafili przenieść pojęcie na nowy materiał (Maruszewski 2003: 303).

Celem badania jest wytworzenie u badanego pojęcia, a więc reprezentacji klipca. W największym skrócie: reprezentacja ta pojawia się wtedy, gdy badany odgadnie, jakie cechy przysługują danemu pojęciu (abstrakcja pozytywna), a jakie mu nie przysługują (abstrakcja negatywna). Testem sprawdzającym znajomość reprezentacji klipca jest umiejętność odróżnienia klipców od nie-klipców.

Można mieć wątpliwości, czy kategoryzacja wymaga pojęć, ale Lewicki zakłada, że tak właśnie jest. Reprezentacja pojawia się u badanego, jeśli potrafi on odróżnić pewne przedmioty. W przykładzie są to figury o określonej barwie i określonym kształcie:

Mianowicie słowo to stanowiło w eksperymencie nazwę pewnej klasy kolorowych obrazków, które wszystkie były podobne do siebie w tym, że stanowiły kwadraty z naklejonym w środku czerwonym, ukośnym kwadracikiem (czerwone karo), a różniły się kolorem tła (Lewicki 1960: 17).

Pojęcie reprezentacji jest składnikiem wyjaśnienia zdolności badanego do rozpoznawania klipców na podstawie cech istotnych tego pojęcia, co stanowi podstawę komunikacji z ludźmi posługującymi się tym pojęciem. Reprezentacja służy więc do wyjaśnienia zachowania oraz komunikacji.