

Problem wyjaśniania neurokognitywnego

6.1. Czym jest wyjaśnienie?

W książce tej bronię poglądu, zgodnie z którym wnioskowanie do najlepszego wyjaśnienia (IBE) odzwierciedla strukturę teorii neurokognitywnych. Idea ta niesie jednak z sobą szereg pytań, z których dwa najistotniejsze brzmią: „jak selekcjonowane są *najlepsze* wyjaśnienia?” oraz „czym w ogóle jest *wyjaśnianie*?”. W poprzednim rozdziale udzieliłem hipotetycznej odpowiedzi na pierwsze z pytań, twierdząc, że najlepsze wyjaśnienia są selekcjonowane za pomocą kryteriów koherencji i konwergencji. Hipotezę tę wzmacniam w kolejnych częściach pracy. W tym rozdziale podejmę natomiast problem samego wyjaśniania.

W rozdziale pierwszym przedstawiłem pogląd, zgodnie z którym „podstawowa neuronauka” wychodzi od pytań „co?” i „jak?”. W przypadku psychologii ewolucyjnej, rozumianej jako paradygmat neuronauki poznawczej, kwestia ta staje się bardziej skomplikowana. Psychologia ewolucyjna odpowiada przede wszystkim na pytanie „dlaczego?”. Zdaniem Ernsta Mayra, wszystkie spośród nauk biologicznych próbują odpowiedzieć na pytanie „co?”. Motywem jest bowiem precyzyjny opis określonych zjawisk¹. Opis jest jednak tylko pierwszym krokiem. W drugim kroku różne dyscypliny należące do nauk biologicznych stawiają i próbują odpowiedzieć na jedno

¹ Za: A. Revonsuo, *O naturze wyjaśniania w neuronaukach*, op. cit., s. 275.

z dwóch pytań. Pierwsze z pytań brzmi „jak?”. Odpowiedź na to pytanie jest zadaniem „podstawowej” neuronauki, podobnie jak np. biologii funkcjonalnej czy też cytologii. Drugie z pytań to „dlaczego?”. Oczywiście jest to domena nauk ewolucyjnych. W teoriach neurokognitywnych pojawia się jednak dodatkowa trudność: zazwyczaj trudno jest od siebie oddzielić wszystkie te pytania.

Analizę problematyki wyjaśniania zacznę od przedstawienia dwóch warunków postawionych przez Carla Hempela. Spełniać musi je każde wyjaśnienie:

(H1) *Warunek istotności*: „informacja, którą wyjaśnienie zawiera, dostarcza racjonalnej podstawy do wiary, że zjawisko wyjaśniane rzeczywiście wystąpiło lub występuje”².

(H2) *Warunek sprawdzalności*: „twierdzenia składające się na wyjaśnienia naukowe muszą podlegać testom empirycznym”³.

Jak okaże się dalej, nie są to jednak jedyne warunki stawiane wyjaśnianiu. Prócz formułowania warunków odpowiedzi trzeba na pytania, takie jak: „czym jest wyjaśnianie?” oraz „w jaki sposób należy wyjaśniać zjawiska?”. Na pytania te odpowiadają poszczególne koncepcje wyjaśniania. W literaturze są one określane mianem *teorii*, *modeli* czy też *strategii*. Ostatnie z tych określeń wydaje mi się najbardziej odpowiednie. Pisał będę też po prostu o *koncepcjach* wyjaśniania. Choć wszystkie ze strategii, które poddam analizie, mają ten sam punkt odniesienia (co, mam nadzieję, stanie się jasne w dalszych częściach tego rozdziału), na poziomie językowym sprawiają one wrażenie zupełnie odmiennych. Jeszcze jedna uwaga terminologiczna: przez *explanandum* rozumiem „to, co podlega wyjaśnieniu”, zaś przez *explanans* – „to, co wyjaśniamy”, czy też „to, za pomocą czego wyjaśniamy”.

Chciałbym również dodać, że nie pretenduję do stworzenia jakiejś nowatorskiej strategii wyjaśniania ani ostatecznego rozstrzyg-

² C.G. Hempel, *Filozofia nauk przyrodniczych*, op. cit., s. 102.

³ *Ibidem*, s. 103.

nięcia problemów związanych z wyjaśnianiem. Postaram się jednak rzucić na nie światło w odniesieniu do neuronauki poznawczej. Moim zdaniem krytyka neopozytywistycznej koncepcji wyjaśniania, która najczęściej owocuje nowymi strategiami wyjaśniania naukowego, nie przekreśla sensowności samych strategii syntaktycznych (zdaniowych). Zwolennicy dominującego w filozofii i metodologii neuronauk poglądu twierdzą, że adekwatną strategią wyjaśniania jest mechanicyzm (wyjaśnianie mechanistyczne lub przyczynowo-mechanistyczne)⁴. Choć, jak pokażę, strategia ta jest *owocna*, trudno uznać ją za całkowicie zadowalającą czy też fundamentalną. Jej siła tkwi raczej w ekonomii opisu.

Zwykle punktem wyjścia w dyskusjach nad wyjaśnianiem naukowym (w ogóle) jest analiza neopozytywistycznej koncepcji Carla Hempla i Paula Oppenheima⁵. Zgodnie z tą strategią wyjaśnienie pewnego zjawiska polega na przeprowadzeniu rozumowania, w którym jedną z przesłanek jest prawo (nauki). Podobnie jak inni autorzy rozpocznę od tej strategii. Najpierw chciałbym jednak wprowadzić pewną dystynkcję. Moim zdaniem różne strategie wyjaśniania można podzielić na trzy grupy. Dlatego też w dalszym ciągu rozdziału zajmę się kolejno strategiami: (1) *syntaktycznymi*, (2) *semantycznymi* oraz (3) *psychologicznymi*. W największym skrócie: na gruncie tych koncepcji wyjaśnienia naukowe są kolejno: *rozumowaniami*, *opisami istniejących stanów rzeczy* oraz *stanami umysłu*. Każdą grupę omówię, kładąc nacisk na najbardziej reprezentatywną koncepcję. Dla grupy (1) będzie to *dedukcyjno-nomologiczna* (neopozytywistyczna) koncepcja wyjaśniania, dla grupy (2) koncepcja *mechanistyczna*, zaś dla grupy (3) koncepcja *reprezentacyjna*.

⁴ Zob. np. C.F. Craver, *Explaining the Brain...*, *op. cit.*; W. Bechtel, *Mental Mechanisms...*, *op. cit.*; W. Bechtel, R.C. Richardson, *Discovering Complexity...*, *op. cit.*; oraz inne publikacje cytowane w dalszej części niniejszej pracy.

⁵ Zob. np. W.C. Salmon, *Four Decades of Scientific Explanation*, University of Pittsburgh Press, Pittsburgh 2006, s. 12 n.; A. Grobler, *Metodologia nauk*, *op. cit.*, s. 103 n. W ostatniej z prac można znaleźć treściwe omówienie ważniejszych koncepcji wyjaśniania. Wykorzystuję je w dalszej części rozdziału.

Do powyższego podziału można jednak zgłosić zastrzeżenia. Przykładowo niektórzy teoretycy mogą domagać się wprowadzenia czwartej grupy, która obejmowałaby *przyczynowe* koncepcje wyjaśniania. Najważniejsza idea tych koncepcji głosi, że *wyjaśnianie polega na identyfikacji przyczyn zjawiska*. Zabieg taki uważam jednak za sztuczny. Pokażę, że nie ma potrzeby rozszerzania podziału o koncepcje przyczynowe. Tym samym usprawiedliwię, że zaproponowany podział spełnia swoje zadanie przynajmniej w odniesieniu do neuro nauki poznawczej.

6.2. Syntaktyczne strategie wyjaśniania

Wszystkie koncepcje wyjaśniania, które można określić mianem syntaktycznych, łączy wspólna cecha. Mówią one, że wyjaśnianie naukowe polega na przeprowadzeniu wniosku. Ujęcie takie jest określane przez Wesleya C. Salmona mianem „trzeciego dogmatu empiryzmu”⁶. Pogląd taki przyjmowali filozofowie nauki, tacy jak Ernest Nagel, Karl Popper czy też wspomniani Carl Hempel i Paul Oppenheim. Najbardziej prominentną i prototypową (określaną nawet niekiedy jako „klasyczna”) koncepcję sformułowali dwaj ostatni⁷. Nosi ona nazwę koncepcji *dedukcyjno-nomologicznej* (przyjmijmy skrót: D-N) lub *modelu podciągania pod prawa* (*covering-law model*). Przyjrzyjmy się tej koncepcji bardziej szczegółowo.

Zgodnie ze strategią D-N wyjaśnianie polega na przeprowadzeniu rozumowania o następujących cechach⁸:

⁶ W.C. Salmon, *Causality and Explanation*, Oxford University Press, Oxford 1998, s. 95 n.

⁷ C.G. Hempel, P. Oppenheim, *Studies in the Logic of Explanation*, „Philosophy of Science” 1948, no. 15, s. 135–175; C.G. Hempel, *Filozofia nauk przyrodniczych*, *op. cit.*, rozdz. 5: *Prawa i ich rola w wyjaśnianiu naukowym*.

⁸ Zob. A. Grobler, *Metodologia nauk*, *op. cit.*, s. 104.

- (1) przesłanki rozumowania to *explanans* (zdania, za pomocą których wyjaśniamy);
- (2) wniosek to *explanandum* (zdanie odnoszące się do wyjaśnianego zjawiska);
- (3) w zbiorze przesłanek (*explanans*) powinno się znajdować minimum jedno prawo;
- (4) przesłanki powinny mieć treść empiryczną.

Zauważmy, że o tym, czy mamy do czynienia z wyjaśnieniem czy nie, decyduje przede wszystkim forma. Dlatego też D-N jest modelem *prima facie* syntaktycznym. Schematycznie wyjaśnianie naukowe wygląda następująco⁹:

$$\frac{L_1, L_2, \dots, L_r}{C_1, C_2, \dots, C_k} \\ E$$

Gdzie L_1, L_2, \dots, L_r , tam prawa (zdania ogólne); gdzie C_1, C_2, \dots, C_k , tam zdania o warunkach początkowych (faktach szczegółowych); E odnosi się zaś do wyjaśnianego zjawiska. Oczywiście strategia D-N została stworzona przede wszystkim z myślą o fizyce i tam najlepiej się sprawdza. Wyjaśniane mogą być nie tylko jednostkowe fakty, ale także regularności oraz prawa (np. Keplera). W przypadku tych ostatnich przesłankami rozumowania są prawa o ogólniejszym zasięgu (np. prawa Newtona). Dodajmy jeszcze, że zbiór przesłanek L_1, L_2, \dots, L_r nie musi odnosić się tylko do zdań, którym skłonni jesteśmy przyznawać specjalny status: „praw naukowych” czy też „zasad przyrody”. Istotna jest jedynie *forma* – muszą być to zdania ogólne (uniwersalne), czyli o postaci: $(\forall x) [F(x) \rightarrow G(x)]$. Jak pisze Hempel:

⁹ Zob. C.G. Hempel, *Filozofia nauk przyrodniczych*, *op. cit.*, s. 107.

Prawa, z których korzysta się w wyjaśnieniach dedukcyjno-nomologicznych, charakteryzują się pewną podstawową wspólną cechą: mają one (...) formę zdań ogólnych. Z grubsza mówiąc, zdanie tego rodzaju stwierdza zachodzenie pewnego regularnie występującego związku między różnymi zjawiskami empirycznymi lub między różnymi aspektami tego samego zjawiska empirycznego. Jest to twierdzenie głoszące, że zawsze i wszędzie, gdzie występują określone warunki F , tam bez wyjątku występują również warunki pewnego innego rodzaju G ¹⁰.

Można podnieść wątpliwość co do stosowalności strategii D-N w *neuroscience* i ogólnie w naukach biologicznych. Nauki biologiczne, w tym *neuroscience*, nie są „modelowymi” naukami nomotetycznymi, a więc takimi, których celem jest formułowanie praw¹¹. Oczywiście nie jest prawdą, że nauki biologiczne w ogóle nie formułują praw. Przykładem są choćby prawa genetyki Mendla czy też prawo Hardy’ego-Weinberga, określające częstość występowania w populacjach genów i alleli¹². Zdaniem Roberta Mayra, ostatnie z nich odgrywa w biologii równie ważną rolę jak pierwsza zasada dynamiki Newtona w fizyce¹³. Zostawmy jednak ten problem. Nawet jeśli „semantycznie” zrezygnujemy z wymogu, aby w wyjaśnianiu korzystać z przesłanek w postaci „praw nauki” czy też „zasad przyrody”, problem pozostaje, gdyż celem nauk biologicznych nie zawsze jest formułowanie zdań ogólnych. Hempel wprowadził jednak jeszcze dwie inne, słabsze strategie wyjaśniania. Mowa tu

¹⁰ *Ibidem*, s. 113.

¹¹ Dokładna analiza tego zagadnienia przekracza ramy niniejszej pracy, dlatego też jestem zmuszony odesłać czytelnika do publikacji, takich jak: M. Lange, *Laws and Theories*, [w:] *A Companion to the Philosophy of Biology*, red. S. Sarkar, A. Plutynski, Blackwell Publishing, Malden, MA 2008, s. 489–506; oraz J. Press, *Physical Explanations and Biological Explanations, Empirical Laws and a priori Laws*, „*Biology & Philosophy*” 2009, no. 24, s. 359–374.

¹² Zob. np. L. Stone, P.F. Lurquin, *Geny, kultura i ewolucja człowieka...*, *op. cit.*, dodatek 2, s. 353–355; A. Łomnicki, *Ekologia ewolucyjna*, *op. cit.*, dodatek 1, s. 187–191.

¹³ Informację podaję za: M.A. Nowak (z udziałem R. Highfielda), *SuperCooperators...*, *op. cit.*, lokalizacja 63.

o koncepcji dedukcyjno-statystycznej (D-S) oraz indukcyjno-statystycznej (I-S). Jeśli chodzi o koncepcję D-S, *explanandum* odnosi się do prawidłowości statystycznej. Wówczas w przesłankach rozumowania powinno się znajdować przynajmniej jedno prawo statystyczne. W przypadku I-S jedną z przesłanek rozumowania powinno być również prawo statystyczne, jednak *explanandum* jest zdarzeniem jednostkowym, o *wysokim prawdopodobieństwie wystąpienia*¹⁴.

Jeśli chodzi natomiast o poziom procesów poznawczych, sytuacja z prawami wygląda jeszcze gorzej. Wielu psychologów uznaje bowiem, że jednym z nielicznych, jeśli nie jedynym, jest przywoływane już przy okazji omawiania ANS prawo Webera-Fechnera. Mówi ono o relacji pomiędzy intensywnością fizycznego bodźca oddziałującego na zmysły (np. intensywność oświetlenia) a siłą reakcji systemu poznawczego człowieka (wzrost zauważalnej intensywności)¹⁵. Prawo to jest wyrażone w formie matematycznej, jednak łatwo można je wypowiedzieć w języku naturalnym: postuluje ono logarytmiczną zależność między czynnikiem oddziałującym na zmysły a reakcją systemu poznawczego. Przeświadczenie niektórych uczonych o wyjątkowym statusie prawa Webera-Fechnera nie oznacza jednak, że w różnych paradygmatach neuronauki poznawczej nie można znaleźć zdań ogólnych, które mogą być wykorzystywane jako przesłanki we wnioskowaniu dedukcyjno-nomologicznym.

Moim zdaniem neopozytywistyczna strategia wyjaśniania naukowego z powodzeniem może być stosowana choćby w przypadku psychologii ewolucyjnej, która jest jednym z podstawowych paradygmatów neuronauki poznawczej. Dla przykładu omówię wnioski, jakie Leda Cosmides oraz John Tooby wyprowadzają z wyników eksperymentów¹⁶ opartych na zadaniu selekcyjnym Wasona (*Wason*

¹⁴ Zob. A. Grobler, *Metodologia nauk, op. cit.*, s. 104–105.

¹⁵ K.J. Holyoak, *Psychology, op. cit.*, s. xii.

¹⁶ Zob. J. Tooby, L. Cosmides, *Cognitive Adaptations for Social Exchange, op. cit.*, s. 163–228.

selection task)¹⁷. Przy okazji można się przyjrzeć samym badaniom oraz sposobowi, w jaki psychologowie ewolucyjni interpretują otrzymane dane.

Uczestnikom zadania Wasona przedstawia się najpierw następującą regułę R_A : *jeśli jedna strona karty zawiera symbol oznaczający samogłoskę, wówczas druga strona karty zawiera symbol oznaczający liczbę parzystą*. Następnie uczestnicy otrzymują cztery karty, na których widnieją kolejno symbole: A, B, 2, 3. Zadaniem uczestnika jest udzielenie odpowiedzi na pytanie: *jakie karty (spośród otrzymanych) należy odwrócić, aby sprawdzić prawdziwość reguły R_A ?* Wyniki wielu eksperymentów wskazują, że większość badanych ma spory problem z poprawną odpowiedzią na to pytanie. Większość z nich skłonna była udzielić następującej odpowiedzi: *należy odwrócić karty z symbolami 2 i B*. Odpowiedź ta jest błędna. W przypadku karty z symbolem 2: reguła nie mówi, że jeżeli karta ma z jednej strony symbol oznaczający liczbę parzystą, wówczas symbol z drugiej strony oznacza samogłoskę. Jeśli chodzi natomiast o kartę z symbolem B: reguła milczy na temat tego, co się dzieje, gdy na jednej stronie karty znajduje się symbol oznaczający spółgłoskę. Poprawna odpowiedź na postawione pytanie jest następująca: *aby sprawdzić prawdziwość reguły R_A , należy odwrócić karty przedstawiające symbole A i 3*. Z zadaniem radzi sobie jedynie od 5% do 30% badanych osób¹⁸. Powyższy przykład będziemy określać jako abstrakcyjną wersję zadania Wasona (Z_A).

Jeśli zamiast kart zawierających abstrakcyjne symbole przedstawi się uczestnikom dane odnoszące się do sytuacji w kontekście społecznym (dlatego będziemy mówić o społecznej wersji zadania Wasona (Z_S)), poprawność udzielonych odpowiedzi znacząco wzrasta. Wówczas zadanie selekcyjne w zasadzie nie sprawia trudności prawie nikomu. Przyjmijmy, że uczestnikom przedstawiono taką oto regułę R_S :

¹⁷ Zob. P.C. Wason, *Reasoning about a Rule*, „Quarterly Journal of Experimental Psychology” 1968, no. A20, s. 273–281.

¹⁸ Zob. M.S. Gazzaniga, *Istota człowieczeństwa...*, op. cit., s. 107–108.

aby dokonać zakupu alkoholu, trzeba mieć ukończone 18 lat. Tym razem zadanie jest następujące: *którego z kupujących należy wylegitymować, aby sprawdzić, czy reguła R_S nie została złamana: kupującego wina, kupującego sok, czterdziestolatka czy szesnastolatka?* Znakomita większość badanych odpowiada na to pytanie poprawnie: *należy wylegitymować kupującego wina oraz szesnastolatka.*

Co ciekawe, zdaniem Cosmides i Tooby'ego, struktura obydwu zadań, czyli Z_S oraz Z_A , jest identyczna. Jediną różnicę stanowią podstawowe dane. Skoro tak, wyjaśnienia domaga się zróżnicowanie poprawności odpowiedzi. Dlaczego badane osoby znacznie lepiej radzą sobie z zadaniem, gdy reguła odnosi się do sytuacji w kontekście społecznym niż do abstrakcyjnych symboli? Cosmides i Tooby tłumaczą tę różnicę, odwołując się do modularnej teorii umysłu. Teoria ta zakłada między innymi, że w rozwiązywanie różnych zadań poznawczych są zaangażowane konkretne, wyspecjalizowane moduły. Sytuacje odwołujące się do reguł relewantnych społecznie aktywują odpowiedni mechanizm, różny od tego, który jest stosowany w przypadku rozumowań abstrakcyjnych. Zdaniem psychologów ewolucyjnych, wydawane sądy oraz decyzje podejmowane w kontekście społecznym powinny cechować się szybkością i niezawodnością. Uzasadniając ten pogląd, psychologowie ewolucyjni twierdzą, że dbanie o sieć relacji społecznych było istotniejsze dla przetrwania i reprodukcji w trudnych plejstocénskich warunkach niż rozwinięcie zdolności radzenia sobie z abstrakcyjnymi łamigłówkami. Relacje społeczne są wyrażane w formie różnego rodzaju reguł, które w wielu przypadkach mogą być sankcjonowane prawnie lub religijnie. Cosmides i Tooby zakładają, że w toku filogenezy zostało wykształcone odpowiednie rozwiązanie (adaptacja) ułatwiające wykrywanie osób, które łamią akceptowane społecznie reguły, takie jak R_S . Rozwiązaniem tym jest działający względnie niezawodnie *moduł poznawczy odpowiedzialny za wykrywanie oszustów*. Jest on uruchamiany w sytuacjach, gdy może być złamana określona norma czy też reguła społeczna, dlatego też nazywany jest *modułem normatywnym*. Działanie tego modułu powinno być uniwersalne, tj. takie samo u wszystkich

ludzi niezależnie od regionu geograficznego i kontekstu kulturowego. Cosmides i Tooby uważają, że warunek ten jest spełniony. Opierają się przy tym na badaniach wykorzystujących różne wersje zadania Wasona, przeprowadzone w różnych kulturach (nawet wśród Indian Shiwiar w Ekwadorze).

Jak pamiętamy, inną istotną cechą modułów jest ich *wrodzoność*. Eksperymenty oparte na zadaniu Wasona były zwykle prowadzone na osobach dorosłych. Trudno wykazać, że moduł normatywny jest wrodzony. Teza ta uległaby jednak wzmocnieniu, gdyby opisane wyżej wyniki były zbieżne z testami przeprowadzonymi na dzieciach. Testy takie przeprowadzili między innymi Paul L. Harris i Maria Núñez¹⁹. W eksperymencie brały udział dzieci trzy- i czteroletnie, które podzielono na dwie grupy. Każdej opowiadano jeden z dwóch typów historyjek odnoszących się do tego samego przedmiotu. W historyjkach pierwszego rodzaju była łamana określona reguła, zaś historyjki drugiego rodzaju zawierały sam opis, nieodwołujący się do reguł. Innymi słowy, historyjki miały albo charakter normatywny (jedna grupa dzieci), albo deskryptywny (druga grupa dzieci). Dzieci z pierwszej grupy usłyszały następującą historyjkę: *Karolina chciałaby coś namalować. Słyszy od mamy, że jeśli chce malować, wówczas powinna ubrać się w fartuszek. Z kolei w grupie drugiej historyjka była taka: Karolina mówi, że jeśli chce coś namalować, wówczas zawsze ubiera fartuszek. W kolejnym etapie dzieciom z obydwu grup zaprezentowano cztery obrazki, przedstawiające kolejno: Karolinę malującą w fartuszku, Karolinę malującą bez fartuszka, Karolinę w fartuszku podczas czynności innej niż malowanie oraz Karolinę bez fartuszka podczas czynności innej niż malowanie. Zadaniem dzieci z grupy, której opowiedziano historyjkę o charakterze normatywnym, było wskazanie obrazka przedstawiającego Karolinę, która jest niegrzeczna i nie spełnia prośby mamy. Dzieci z drugiej grupy miały natomiast pokazać obrazek przedstawiający Karolinę, która robi coś in-*

¹⁹ Zob. P.L. Harris, M. Núñez, *Understanding of Permission Rules by Preschool Children*, „Child Development” 1996, no. 67, s. 1572–1591.

nego niż deklarowała. Wyniki eksperymentu były następujące: dzieci z grupy normatywnej wybierały poprawny obrazek znacznie częściej (72% trzylatków oraz 83% czterolatek) niż dzieci z grupy, w której opowiedziano historyjkę deskryptywną (jedynie 40% – zarówno trzylatków, jak i czterolatek). Badacze stwierdzili, że rozumowania o charakterze normatywnym cechują się większą trafnością niż rozumowania deskryptywne. Teza ta jest oczywiście spójna z podstawowymi wynikami Cosmides i Tooby'ego.

Powstanie modułu wykrywania oszustów (normatywnego) wpisuje się w scenariusz ewolucyjny, przewidujący istnienie „wyścigu zbrojeń” pomiędzy oszustami a oszukanymi. Zwolennicy tzw. hipotezy makiawelicznej inteligencji (*Machiavellian intelligence hypothesis*) twierdzą, że wynikiem tego wyścigu był mający miejsce około dwóch milionów lat temu szybki rozrost kory mózgowej, który umożliwił rozwój inteligencji²⁰. Hipoteza ta zakłada ponadto związek między zdolnościami społecznymi, które wymagają często tytułowego „makiawelizmu”, a zdolnościami kognitywnymi²¹. Robert Trivers pisze, że:

Oszust i oszukiwany są uwięzieni w koewolucyjnej walce, której wynikiem jest ciągle ulepszanie adaptacji po obydwu stronach konfliktu. Jedną z takich adaptacji jest inteligencja. Dowody są jasne i przytłaczające: zarówno wykrywanie oszustw, jak i ich propagowanie były głównymi siłami wspierającymi ewolucję inteligencji²².

Nikt nie lubi być oszukiwany, zarówno w drobnych, jak i istotnych kwestiach. Za wszelką cenę próbujemy ustrzec się przed oszustwem, szczególnie jeśli dotyczy ono ważnych spraw, takich jak

²⁰ Zob. A. Whiten, *Machiavellian Intelligence Hypothesis*, *op. cit.*, s. 495–496.

²¹ Więcej na temat związków między oszukiwaniem a zdolnościami poznawczymi można znaleźć w pracach: M. Hohol, *Zjawisko kłamstwa w perspektywie nauk neurokognitywnych i ewolucyjnych*, *op. cit.*; M. Hohol, P. Urbańczyk, *Self-Deception. Between Philosophy and Cognitive Neuroscience*, *op. cit.*

²² R. Trivers, *The Folly of Fools. The Logic of Deceit and Self-Deception in Human Life*, *op. cit.*, e-book edition, lokalizacja 26.

przetrawianie czy reprodukcja. Prócz samej inteligencji owocem wysiłku zbrojeń pomiędzy oszustami i oszukiwanymi mogło być wykształcenie modułu wykrywania oszustów. Wróćmy jednak do strategii wyjaśniania, którą stosują psychologowie ewolucyjni.

Teoria głosząca istnienie modułu wyspecjalizowanego w wykrywaniu oszustów (normatywnego) jest wspierana również przez badania lezji. W wyniku wypadku, którego skutkiem było uszkodzenie mózgu, zdolność pacjenta o inicjałach R.M. do przeprowadzania rozumowań w kontekście społecznym została upośledzona. Z kolei zdolności do rozumowań w kontekstach nierelevantnych społecznie nie zostały naruszone²³. Innymi słowy, R.M. w poprawny sposób przeprowadzał operacje o charakterze deskryptywnym, natomiast stracił zdolność do rozumowań normatywnych. Zdaniem Cosmides, Tooby'ego i ich współpracowników, świadczy to o niezależności obydwu typów rozumowań i stojących za nimi systemów poznawczych.

Psychologowie ewolucyjni sądzą, że konkretne wyzwania, na jakie napotyka organizm, uruchamiają odpowiednie dla nich moduły. W rozważanym przypadku detekcja złamania reguły R_S jest możliwa dzięki funkcjonowaniu modułu wykrywania oszustów. Jedną z cech modułów jest ich specjalizacja. Dlatego też moduł wykrywania oszustów nie jest uruchamiany w przypadkach nierelevantnych społecznie, takich jak testowanie poprawności reguły R_A .

Można powiedzieć, że istnieje prawidłowość, którą można zapisać jako zdanie uniwersalne, czyli o postaci $(\forall x) [F(x) \rightarrow G(x)]$. Zdanie to charakteryzuje działanie modułu wykrywania oszustów. *Za każdym razem – ceteris paribus – gdy organizm znajdzie się w sytuacji możliwego złamania reguły społecznej, uruchamiany jest działający sprawnie moduł wykrywania oszustów.* Nic nie stoi więc na przeszkodzie, by zdanie to potraktować jako prawo, które może być wykorzystane w wyjaśnieniu dedukcyjno-nomologicznym. Wyja-

²³ V.E. Stone, L. Cosmides, J. Tooby, N. Kroll, R.T. Knight, *Selective Impairment of Reasoning about Social Exchange in a Patient with Bilateral Limbic System Damage*, „Proceedings of the National Academy of Sciences” 2002, no. 99, s. 11531–11536.

śnienie dużej poprawności w rozwiązywaniu społecznej wersji zadania Wasona przebiega według znanego nam już schematu:

L: Za każdym razem, gdy organizm znajdzie się w sytuacji możliwego złamania reguły społecznej, uruchamiany jest działający sprawnie moduł wykrywania oszustów.

C: Badana osoba znalazła się w sytuacji możliwego złamania reguły R_s : aby dokonać zakupu alkoholu, trzeba mieć ukończone 18 lat.

E: Uruchomiony został moduł wykrywania oszustów, co pozwoliło udzielić prawidłowej odpowiedzi, kogo należy skontrolować, aby sprawdzić, czy nie łamie R_s .

Powyższy schemat wyjaśniania można by również zastosować w innych przypadkach, gdy w grę wchodzi działanie modułów, o których piszą psychologowie ewolucyjni. Nawet jeśli tak jest, nie determinuje to jeszcze tego, że neopozytywistyczną strategię wyjaśniania można zastosować do wszystkich zjawisk występujących w ramach neuronauki poznawczej. Psychologia ewolucyjna jest poza tym tylko jednym (na dodatek bardzo kontrowersyjnym) z paradygmatów neuronauki poznawczej. Nie wszystkie paradygmaty dążą jednak za wszelką cenę do formułowania zdań ogólnych na temat procesów poznawczych. Można by zapewne argumentować, że tam, gdzie nie da się wyjaśnić w sposób dedukcyjno-nomologiczny, można zastosować inne warianty, tj. dedukcyjno-statystyczny lub indukcyjno-statystyczny. Nawet biorąc pod uwagę tylko przytoczone przeze mnie informacje na temat prawdopodobnego scenariusza ewolucyjnego, który doprowadził do powstania modułu normatywnego oraz danych związanych z leżjami, można by argumentować, że sformułowane wyżej wyjaśnienie dedukcyjno-nomologiczne jest bardzo powierzchowne: dotyka ono jedynie pytania „jak?”, pomijając natomiast pytanie „dlaczego?”. Ponadto wyjaśnienie to opiera się jedynie na obserwacji psychologiczno-poznawczej regularności L , pomijając rolę

niższych poziomów biologicznych. Co więcej, wielu badaczy twierdzi, że wyjaśnienie skuteczności rozumowań normatywnych zaproponowane przez Cosmides i Tooby'ego jest po prostu błędne.

Dan Sperber, Vittorio Girotto oraz Francesco Cara podważyli tezę Cosmides i Tooby'ego²⁴. W szczególności udało im się sformułować zadania angażujące rozumowania normatywne (w kontekście społecznym), w których błędne odpowiedzi wynosiły ponad 80%. Co więcej, pewne wersje zadań deskryptywnych przyniosły więcej poprawnych odpowiedzi niż w testach Tooby'ego i Cosmides. Zdaniem Sperbera i Girotto, poziom poprawnych odpowiedzi nie jest zależny od tego, czy zadanie jest sformułowane w sposób normatywny czy deskryptywny. Według nich ilość poprawnych rozwiązań danego zadania jest uzależniona od stopnia odpowiedniości pomiędzy okresem warunkowym a zadaniem pytaniem. Poprawność odpowiedzi jest zatem uzależniona od tego, czy badani *rozumieją* zadanie, nie zaś od tego, czy „odpala” ono działanie określonych modułów (jeśli takowe w ogóle istnieją). Najprawdopodobniej wyniki uzyskane przez Tooby'ego i Cosmides były powodowane tym, że badani *nie rozumieli* abstrakcyjnego zadania (Z_A), podczas gdy wiedzieli doskonale „o co chodzi”, gdy zadanie było osadzone w kontekście społecznym (Z_S)²⁵. Można zgłosić jeszcze jedną uwagę krytyczną co do samego sformułowania zadania selekcyjnego Wasona. Otóż wątpliwość może budzić stwierdzenie, że obydwa warianty – abstrakcyjny (z kartami) i społeczny – są równoważne. Uwagę można zwrócić choćby na modalności, jakie pojawiają się w wariacie społecznym, a których brak w wariacie abstrakcyjnym. W moim przekonaniu samo w sobie nie podważa

²⁴ D. Sperber, V. Girotto, *Use of Misuse of the Selection Task? Rejoinder to Fiddick, Cosmides, and Tooby*, „Cognition” 2002, no. 85, s. 277–290; D. Sperber, F. Cara, V. Girotto, *Relevance Theory Explains the Selection Task*, „Cognition” 1995, no. 57, s. 31–95.

²⁵ Problem istnienia modułu normatywnego oraz zagadnienia związane z zadaniem selekcyjnym Wasona zostały szczegółowo omówione w pracy: B. Kucharzyk, *Is There a Normative Module? Some Remarks on the Wason Selection Task Experiments in the Field of Normative Reasoning*, [w:] *The Many Faces of Normativity*, *op. cit.*

to całkowicie tezy o modularnym charakterze umysłu, jednak znacząco osłabia argumentację Tooby'ego i Cosmides.

Wróćmy jednak na poziom bardziej ogólnych analiz metodologicznych. Podstawowy problem z wyjaśnianiem w stylu neopozytywistycznym polega na tym, że wszystkie trzy rodzaje, czyli D-N, D-S, I-S, generują paradoksy uwidaczniające różne niedostatki tych koncepcji. Paradoksy te stały się impulsem do tworzenia alternatywnych strategii wyjaśniania. Przyjrzyjmy się pokrótce niektórym z paradoksów²⁶, zaczynając od problemów, w jakie jest uwikłane wyjaśnianie dedukcyjno-nomologiczne.

Pierwszy paradoks dotyczy *masztu i rzuconego przez niego cienia*. Prawdopodobnie po raz pierwszy został on zauważony przez Sylvaina Brombergera²⁷. Paradoks ten ma pokazać, że wyjaśnianie dedukcyjno-nomologiczne nie spełnia tzw. *warunku asymetrii*. Zgodnie z tym warunkiem, jeśli zjawisko (prawidłowość) *A* można wyjaśnić, odwołując się do zjawiska (prawidłowości) *B*, to nie powinno być możliwe wyjaśnienie *B* poprzez odwołanie się do *A*. Długość cienia, jaką rzuca maszt, można wyjaśnić, przeprowadzając rozumowanie, którego przesłankami są odpowiednie prawa (optyki i trygonometrii) oraz fakty jednostkowe (wysokość masztu i wartość kąta). Analogicznie można również wyjaśnić długość cienia. Jedyna różnica polega na tym, że zamiast wysokości masztu za przesłankę bierzemy długość cienia. Paradoks polega na tym, że koncepcja D-N dopuszcza możliwość wyjaśnienia wysokości masztu, mając za przesłankę długość cienia.

Barometr i burza to kolejny paradoks, sformułowany tym razem przez Hansa Reichenbacha. Uwidacznia on problem korelacji i oddziaływań przyczynowych. Zdanie „nadchodzi burza” można

²⁶ Wszystkie paradoksy wyjaśniania referuję za Adamem Groblerem: *idem, Metodologia nauk, op. cit.*, s. 105–112; *idem, Prawda i racjonalność naukowa, op. cit.*, s. 157–162. Omówienie większości z nich można znaleźć praktycznie w dowolnym podręczniku do filozofii nauki czy metodologii nauk.

²⁷ S. Bromberger, *Why Questions*, [w:] *Mind and Cosmos. Essays in Contemporary Science and Philosophy*, red. R.G. Colodny, University of Pittsburgh Press, Pittsburgh 1966.

wywnioskować, mając za przesłanki zdanie ogólne „Zawsze, gdy wskazówka barometru opada, nadchodzi burza” oraz zdanie o warunkach początkowych: „Wskazówka barometru opadła”. Trudno zaakceptować wyjaśnienie zjawiska atmosferycznego poprzez odwołanie się do obserwacji wskazówki urządzenia. Znacznie rozsądniej wyjaśnić zarówno nadejście burzy, jak i zachowanie wskazówki spadkiem ciśnienia atmosferycznego (jest to wspólna przyczyna obydwu faktów).

Inny paradoks jest związany z wyjaśnianiem zjawisk takich jak *zaćmienie Słońca*. Paradoks ten uzmysławia, że wyjaśnienia dedukcyjno-nomologiczne nie rozróżniają należycie predykcji i retrodykcji. Zaćmienie Słońca można wyjaśnić poprzez predykcję tego zjawiska. Należy przeprowadzić rozumowanie, którego przesłankami są prawa ruchu ciał niebieskich oraz zdania odnoszące się do konfiguracji Słońca, Księżyca i Ziemi. Problem polega na tym, że analogicznie można zdobyć wiedzę o przeszłych zaćmieniach. Wówczas *przeszłe* zaćmienie byłoby wyjaśniane *aktualną* konfiguracją ciał niebieskich, co wydaje się niedorzeczne. Dwa ostatnie paradoksy pokazują, że niektóre, ale nie wszystkie, predykcje mogą być jednocześnie wyjaśnieniami, zaś retrodykcje nigdy.

Ostatni paradoks modelu D-N, o którym chcę wspomnieć, nosi nazwę paradoksu *mężczyzny i pigułki*. Zdanie „Jaś nie zajdzie w ciążę” może być wyprowadzone z prawa „osoby zażywające pigułki antykonceptyjne – *ceteris paribus* – nie zachodzą w ciążę” oraz zdania jednostkowego „Jaś zażywa pigułki”. Trudno jednak uznać powyższe rozumowanie za wyjaśnienie. Paradoks ten pokazuje, że wyjaśnienia muszą opierać się na czynnikach relewantnych przyczynowo (to, że Jaś jest mężczyzną, jest wystarczającą przyczyną, by nie mógł zająć w ciążę; wszystkie inne dodatkowe czynniki są już nieistotne).

Przejdźmy teraz do paradoksów, jakie generują słabsze, czyli statystyczne wersje wyjaśniania w stylu neopozytywistycznym. Paradoks *kataru i witaminy C* pokazuje kolejne problemy wyjaśniania związane z przyczynowością. Załóżmy, że Małgosia miała katar, który ustąpił, gdy zaczęła zażywać witaminę C. Przesłankami rozu-

mowania, które ma dostarczyć wyjaśnienia, są: prawo statystyczne mówiące o leczniczych własnościach witaminy C oraz zdanie mówiące o zażyciu witaminy C przez Małgosię. Problem polega jednak na tym, że – zgodnie z ludowym porzekadłem – często katar leczony witaminą C trwa siedem dni, a nieleczony tydzień. Rozumowanie mówi tylko, że prawdopodobieństwo wyzdrowienia Małgosi było wysokie. To za mało, by mówić o wyjaśnianiu naukowym. Trzeba by jeszcze wykazać – na zasadzie kontrastu – że prawdopodobieństwo wyzdrowienia bez zażywania witaminy C jest istotnie niższe. Na bardzo podobnej zasadzie funkcjonuje podany przez Wesleya Salmona paradoks *psychoterapii*. W procesie wyjaśniania nie wystarczy wskazać, że prawdopodobieństwo ustania dolegliwości psychicznych jest wysokie w przypadku zastosowania psychoterapii. Należy jeszcze wskazać, że jest ono istotnie wyższe niż prawdopodobieństwo samoczynnej poprawy zdrowia psychicznego.

Inny paradoks uzmysławiający problemy modelu I-S to *kiła i pareza*. Pareza, zwana również paraliżem postępowym, występuje u około 10% osób z nieleczoną kiłą pierwszorzędową. Oczywistym wyjaśnieniem zachorowania Jasia na paręzę jest to, że wcześniej cierpiał na kiłę. Jak pamiętamy, Hemplowska koncepcja I-S wymaga, by prawdopodobieństwo zajścia wyjaśnianego zdarzenia było *wysokie*. Oczywiście jest to kwestia uznaniowa, jednak 10% to intuicyjnie raczej dość niskie prawdopodobieństwo. Paradoks ten świadczy o tym, że niskie prawdopodobieństwo zajścia zjawiska nie powinno przekreślać wyjaśniania statystycznego. Na podobnej zasadzie funkcjonuje paradoks *nieprawidłowej monety*. Przypuśćmy, że gramy monetą skonstruowaną w taki sposób, iż prawdopodobieństwo wyrzucenia orła wynosi aż 95%, a reszki tylko 5%. Wyrzucenie orła nie różni się istotnie od wyrzucenia reszki, dlatego też wyjaśnianie powinno radzić sobie jednakowo z obydwojma zjawiskami. Jednak na gruncie koncepcji indukcyjno-statystycznej wyjaśnić można tylko wyrzucenie orła. Omówione wyżej paradoksy są swego rodzaju drogowskazem. Okazuje się, że Hemplowskie warunki (H1) i (H2) są konieczne, ale nie są wystarczające. Paradoksy prowadzą do sformułowania

dotatkowych warunków, jakie powinno spełniać dobre wyjaśnienie naukowe. Warunki te sprawiają, że pojęcie wyjaśniania naukowego nabiera aspektu normatywnego. Oto one:

(W1) Warunek następstwa czasowego: samo następstwo czasowe zjawisk nie stanowi ich wyjaśnienia.

(W2) Warunek asymetrii: przyczyny powinny wyjaśniać efekty, a nie odwrotnie.

(W3) Warunek wspólnej przyczyny: przyczynowo niezależne efekty wspólnych przyczyn nie wyjaśniają się nawzajem.

(W4) Warunek relewancji: zjawiska nieistotne przyczynowo nie są wyjaśnieniami.

(W5) Warunek prawdopodobieństwa: przyczyny nie muszą czynić efektów wysoce prawdopodobnymi, aby je wyjaśnić (prawdopodobieństwo zajścia wyjaśnianego zdarzenia nie musi być wysokie)²⁸.

Opisane wyżej paradoksy pokazują, że wyjaśnianie w stylu neo-pozytywistycznym nie docenia roli przyczynowości. Opozycyjnym, przynajmniej na poziomie deklaracji, jest pogląd upatrujący istoty wyjaśniania we wskazaniu przyczyny interesującego zjawiska. Przyczynowa koncepcja wyjaśniania nie jest oczywiście niczym nowym. Jej źródeł można szukać już u Arystotelesa, zaś w czasach współczesnych spopularyzował ją Michael Scriven²⁹. Powołanie się na prawa może być pomocne, ale *nie jest* obligatoryjne w procesie wyjaśniania. Aby wyjaśnić zjawisko, jakim jest plama na dywanie, wystarczy zauważyć, że jej *przyczyną* było rozlanie się atramentu. Samo podanie przyczyny jest więc tożsame z wyjaśnieniem.

Oczywiście koncepcja przyczynowa nie jest pozbawiona również trudności. Po pierwsze, wyjaśniając poprzez podanie przyczyny, powinniśmy najpierw jasno wyartykułować, co rozumiemy

²⁸ C.F. Craver, *Explaining the Brain...*, *op. cit.*, s. 26.

²⁹ Zob. M. Scriven, *Explanations, Predictions, and Laws*, [w:] *Scientific Explanation, Space, and Time*, red. H. Feigl, G. Maxwell, University of Minnesota Press, Minneapolis 1962, s. 170–230; zob. A. Grobler, *Metodologia nauk*, *op. cit.*, s. 107.

przez *przyczynowość*. Zwolennicy wyjaśniania przyczynowego nie czynią tego zawsze. Pomińmy jednak ten problem. W kontekście neuronaukowym i ogólnie biologicznym należałoby określić, *jakie* przyczyny są tożsame z wyjaśnieniami. Popularna jest dystynkcja na przyczyny bliższe i dalsze. Prosty przykład wyjaśnienia opar- tego na przyczynie bliższej dotyczy potencjału czynnościowego neuronu. Wyjaśnieniem depolaryzacji błony na poziomie, który pozwala wygenerować impuls nerwowy, jest podanie przyczyny tego zjawiska. Przyczyną tą jest dostarczenie odpowiednio silnego prądu dokomórkowego. O ile pojęcie przyczyny bliższej jest łatwe do zrozumienia, o tyle pojęcie przyczyny dalszej, nawiązujące zwykle do wyjaśniania ewolucyjnego, jest znacznie bardziej mętne. Gdyby każdy wytwór działania doboru naturalnego był adaptacją, wyjaśnianie poprzez odwołanie się do przyczyn dalszych nie byłoby skomplikowane. Jednak, jak powiedziałem w rozdziale drugim, nie wszystkie cechy obiektów biologicznych są adaptacjami. W związku z tym bardzo trudno jest wskazać przyczyny dalsze. Z kolei odwoływanie się jedynie do przyczyn bliższych wydaje się dość powierzchowne. Jest to jednak tylko odsłona ogólniejszej trudności wyjaśniania przyczynowego.

Bardzo rzadko można wskazać *jedną jedyną* przyczynę pewnego zjawiska³⁰. Zdarzenia są bowiem powodowane całymi łańcuchami czy też splotami przyczynowymi. Wyjaśnianie przyczynowe mogłoby ograniczać się jedynie do podawania przyczyn najbardziej istotnych, takich jak w powyższych przykładach. Można tego dokonać, tylko jeśli uda się nam rozróżnić przyczyny bardziej i mniej istotne. Nie wiadomo, czy rozróżnienie takie jest w ogóle możliwe i sensowne. Nie wiadomo tym bardziej, jak można tego dokonać dla złożonych zjawisk.

Przejdźmy jednak do zagadnienia bardziej ogólnego. Dlaczego nie ma sensu wyróżniać koncepcji przyczynowych (M. Scriven) jako osobnej strategii wyjaśniania? Aby odpowiedzieć na to pytanie,

³⁰ Zob. *ibidem*, s. 110–111.

powołałam się na słynny argument sformułowany przez Paula Thagarda³¹. Jego zdaniem wszystkie znane w filozofii nauki koncepcje wyjaśniania zawierają wspólny komponent. Komponentem tym jest *przyczynowość*. Przykładowo, w omówionej wyżej strategii wyjaśniania w stylu neopozytywistycznym rozumowanie wyjaśniające jest pewnym ciągiem przyczyno-skutkowym, którego ostatnim elementem jest *explanandum*. Kategoria przyczynowości pojawia się natomiast *explicite* w strategii semantycznej.

6.3. Semantyczne strategie wyjaśniania

Opisane w poprzednim paragrafie paradoksy nie są jedynymi problemami, w jakie – zdaniem wielu filozofów nauki i metodologów – uwikłane są neopozytywistyczne czy, ogólniej mówiąc, syntaktyczne koncepcje wyjaśniania naukowego. Zwolennicy różnorodnych koncepcji semantycznych zarzucają im wręcz popełnienie fundamentalnej pomyłki co do istoty wyjaśniania. Zwolennikiem strategii semantycznej jest Carl Craver. Jego zdaniem wyjaśnianie powinno odwoływać się do *realnie istniejących* stanów rzeczy, czy też obiektów i ich własności. Twierdzi on, że obiektywne wyjaśnienia „są rodzajami rzeczy, które odkrywamy i opisujemy”³². Pogląd, zgodnie z którym wyjaśnianie polega na przeprowadzaniu rozumowań (operacji na zdaniach), jest według niego pomyłką spowodowaną niezrozumieniem natury wyjaśniania. Wyjaśnianie ma polegać według niego na odkrywaniu *mechanizmów*, które są „obiettami i działaniami zorganizowanymi w taki sposób, że przedstawiają *wyjaśniane zjawisko (explanandum phenomenon)*”³³. Przyjęcie strategii semantycznej prowadzi ponadto do akceptacji wyjaśnień w postaci rysunków, wykresów i diagramów. Bardzo często oddają one bowiem istotę wyjaśnianego zjawiska

³¹ Zob. P. Thagard, *Conceptual Revolutions*, Princeton University Press, Princeton, NJ 1992, s. 118–130.

³² C.F. Craver, *Explaining the Brain...*, *op. cit.*, s. 27.

³³ *Ibidem*, s. 6.

w znacznie większym stopniu niż opisy słowne. Przykładowo, wykorzystanie fMRI powoduje, że mechanizmy są reprezentowane jako trójwymiarowe obiekty³⁴. Podejście takie koresponduje z potocznym przeświadczeniem głoszącym, że metody neuroobrazowania pozwalają tworzyć „obrazy umysłu”. Jeśli zaś chodzi o kontekst metodologiczny, strategia semantyczna koresponduje z ikoniczną koncepcją modelu. Jak pisze Adam Grobler, model ikoniczny to:

Symboliczna reprezentacja badanego zjawiska, która ma charakter podobizny, to znaczy niektóre cechy odwzorowują czy naśladują (z dokładnością do założeń idealizacyjnych) określone cechy oryginału (...). Ten sam oryginał może mieć różne podobizny w zależności od cech, będących przedmiotem zainteresowania³⁵.

Modele ikoniczne są wykorzystywane w niezdaniowych ujęciach teorii naukowych, o których wspomniałem w rozdziale pierwszym. Nie wszystkie modele są jednak zarazem wyjaśnieniami. Jak zauważa Craver, wiele modeli służy jedynie do systematyzacji różnych danych³⁶.

Jak nadmieniałem już wcześniej, poglądy Carla Cravera przynależą do istotnego nurtu metodologii i filozofii neuronauki poznawczej, określanego mianem mechanicyzmu. O popularności tego nurtu świadczą choćby słowa Williama Wimsatta, który pisze, iż: „ostatnio w biologii wielu naukowców zauważa, że ich praca polega na wyjaśnianiu zjawisk poprzez odkrywanie mechanizmów”³⁷. Jeden z największych zwolenników mechanistycznej strategii wyjaśniania naukowego, William Bechtel, uważa, że wiedza z zakresu nauk biologicznych – w tym badających mózg – jest rzadko kiedy przedstawiana

³⁴ A. Revonsuo, *O naturze wyjaśniania w neuronaukach*, op. cit., s. 277–278.

³⁵ A. Grobler, *Metodologia nauk*, op. cit., s. 175–176.

³⁶ Zob. C.F. Craver, *When Mechanistic Models Explain*, „Synthese” 2006, vol. 153, no. 3 (*Neuroscience and Its Philosophy*), s. 355–376.

³⁷ W. Wimsatt, *Complexity and Organization*, [w:] *Proceedings of the 1972 Biennial Meeting of the Philosophy of Science Association*, red. K.F. Schaffner, R.S. Cohen, Reidel, Dordrecht 1972, s. 67.

w formie ogólnych praw³⁸. Jego zdaniem praktyka badawcza pokazuje, że wyjaśnianie zagadkowego zjawiska czy też struktury polega raczej na rozpoznaniu składających się na nie elementów i mikrostruktur, zbadaniu relacji pomiędzy tymi elementami i pokazaniu, jak elementy te generują dane zjawisko (strukturę)³⁹. Według Bechtela odpowiedź na pytanie „jak?” polega więc na rekonstrukcji mechanizmu. W opinii zwolenników mechanicyzmu podejście to współgra dobrze z wielopoziomowością neuronauki poznawczej. Mechanizmy mogą być reprezentowane w sposób adekwatny do specyfiki danego poziomu. Bechtel twierdzi, że wszystkie poziomy teorii neurokognitywnych, w tym także poziom procesów poznawczych, podlegają wyjaśnianiu mechanistycznemu. Dodam jeszcze, że mechanicyzm jest szeroko przyjmowanym poglądem, czy też raczej ramą pojęciową (*theoretical framework*) w ogólnie rozumianej filozofii nauk biologicznych. Wyjaśnianie mechanistyczne można uznać za modelowy przykład strategii semantycznej⁴⁰. W niniejszym paragrafie skupię się na analizie wyjaśniania mechanistycznego w kontekście neuronauki poznawczej.

Sam termin *mechanizm* jest bardzo często wykorzystywany w rozmaitych naukach, zwykle jednak jego znaczenie nie jest określone precyzyjnie. Ma on również bogatą tradycję filozoficzną, której nawet pobieżne omówienie zdecydowanie przekraczałoby ramy niniejszej pracy⁴¹. W kontekście mechanistycznej strategii wyjaśniania w neuronauce poznawczej mechanizm jest pewnym wzorcem działania, który możemy *odkryć* i dzięki któremu możemy *wyjaśnić* in-

³⁸ Zob. W. Bechtel, *Levels of Description and Explanation in Cognitive Science*, „Minds and Machines” 1994, no. 4, s. 1–25.

³⁹ Zob. W. Bechtel, *Mental Mechanisms...*, *op. cit.*, rozdz. 1; W. Bechtel, R.C. Richardson, *Discovering Complexity...*, *op. cit.*

⁴⁰ Tym bardziej różne inne semantyczne koncepcje wyjaśniania, takie jak np. analiza funkcjonalna (zob. R. Cummins, *Functional Analysis*, „The Journal of Philosophy” 1975, vol. 72, no. 20, s. 741–765), bez większego problemu można przedstawić w terminach mechanistycznych.

⁴¹ W związku z tym zainteresowanych czytelników odsyłam do pracy Michała Hellera i Józefa Życińskiego: *Wszechświat – maszyna czy myśl? Filozofia mechanicyzmu: powstanie – rozwój – upadek*, PTT, Kraków 1988.

teresujące *zjawisko*. W ważnym i często cytowanym artykule Peter Machamer, Lindley Darden oraz Carl Craver podają następującą definicję mechanizmów:

Mechanizmy są obiektami [*entities*] [o określonych własnościach – M.H.] i procesami [*activities*] zorganizowanymi w taki sposób, że powodują one regularne zmiany, począwszy od początku, czy też warunków początkowych, aż do zakończenia [działania – M.H.] lub warunków końcowych⁴².

Powyższą definicję przyjmuję jako roboczą w dalszej części tego paragrafu i określam ją skrótowo jako MDC⁴³. Zdaniem jej twórców, definicja MDC spełnia wymagania odnośnie do podstawowego mechanizmu neuronalnego. Jest nim transmisja sygnału z neuronu presynaptycznego do postsynaptycznego, która zachodzi dzięki uwolnieniu molekuł chemicznego neuroprzekaźnika do szczeliny synaptycznej, łączeniu ich przez odpowiednie receptory i depolaryzacji komórki postsynaptycznej. Nieco dalej rozważę ten mechanizm bardziej szczegółowo. Na razie dodam jeszcze, że przytoczona wyżej definicja jest z powodzeniem wykorzystywana także w przypadku wielu innych mechanizmów biologicznych, takich jak np. replikacja DNA⁴⁴.

W definicji MDC była mowa o obiektach (*entities*) i procesach (*activities*). Przyjrzyjmy się bliżej wymienionym pojęciom:

Mechanizmy są złożone zarówno z obiektów (wraz z ich własnościami), jak i procesów. Procesy wywołują zmiany. Obiekty biorą

⁴² P.K. Machamer, L. Darden, C.F. Craver, *Thinking about Mechanisms*, *op. cit.*, s. 3; ten sposób analizy mechanizmów został rozwinięty najbardziej przez Carla Cravera w książce *Explaining the Brain...*, *op. cit.*

⁴³ Nie oznacza to, że MDC jest jedyną i powszechnie akceptowaną definicją mechanizmu. Inna definicja, przyjmowana przez Stuarta Glennana, jest zbudowana na pojęciach takich jak *obiekt*, *interakcja* oraz *prawo przyczynowe*. Z kolei Wesley Salmon za kryterium wyodrębnienia mechanizmu uznaje możliwość transmisji abstrakcyjnych wzorców oddziaływań.

⁴⁴ Zob. P.K. Machamer, L. Darden, C.F. Craver, *Thinking about Mechanisms*, *op. cit.*, s. 3.

udział w zmianach. Procesy zazwyczaj wymagają, by obiekty były wyposażone w określone własności⁴⁵.

Tworzące dany mechanizm, obiekty mają określone *położenie przestrzenne, strukturę oraz hierarchię*. Z kolei procesy charakteryzują się określonym *porządkiem czasowym, tempem oraz czasem trwania*. Mechanizmy działają w *regularny* sposób. Oznacza to, że zazwyczaj w określonych warunkach jest możliwe przewidywanie zmian w układzie nerwowym, jakie wywołuje dany mechanizm. Działanie mechanizmu polega na regularnej transmisji określonego wzorca od warunków początkowych, przez etapy pośrednie, aż do stanu końcowego. Co istotne, transmisja ta jest ciągła. W związku z tym, że podstawowe terminy mechanicyzmu to *obiekt i działanie*, definicja MDC jest określana mianem *dualistycznej*. Jej twórcy przeprowadzają analizę statusu ontycznego mechanizmów. Ich zdaniem ani substancjalizm (głoszący pierwotność obiektów), ani filozofia procesu (mówiąca o redukcji obiektów do procesów) nie są stanowiskami w pełni zadowalającymi⁴⁶. Dzieje się tak, ponieważ obiekty o określonych własnościach mogą uczestniczyć tylko w niektórych działaniach. Z drugiej strony działania determinują to, jakie obiekty mogą w nich uczestniczyć. Innymi słowy, obiekty i działania są powiązane i współzależne⁴⁷. Zauważmy, że – ogólnie rzecz biorąc – poglądy Machamera, Darden i Cravera są osadzone w nurcie realizmu naukowego. Mechanistyczna strategia wyjaśniania naukowego jest więc uwikłana w określone poglądy na temat statusu poznawczego teorii naukowych.

Kolejnym istotnym terminem dla wyjaśniania mechanistycznego jest *dekompozycja*. Dekompozycja polega na wyodrębnieniu obiektów i procesów konstytuujących mechanizm. Inaczej mówiąc, chodzi o wskazanie części, z których składa się interesujący system, określe-

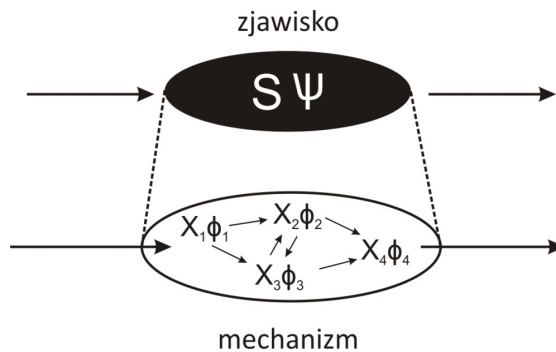
⁴⁵ *Ibidem*.

⁴⁶ Zob. *ibidem*, s. 4–5.

⁴⁷ Zob. *ibidem*, s. 6.

niu, jakie pełnią one funkcje i jakie interakcje zachodzą między nimi. Ponownie warto podkreślić, że proces dekompozycji jest zwykle rozumiany przez zwolenników mechanicyzmu w duchu realizmu naukowego. Ich zdaniem za pomocą dekompozycji można dotrzeć do *rzeczywiście istniejących* obiektów i procesów. Z drugiej strony, zdaniem niektórych filozofów nauki, nawet jeśli metoda ta nie pozwala dotrzeć w danym przypadku do „istoty rzeczy” (np. ze względu na zbyt dużą złożoność systemu), wciąż jest ona użyteczna w analizie eksplanacyjnej⁴⁸.

Po omówieniu znaczenia podstawowych terminów możemy przyjrzeć się schematowi wyjaśniania mechanistycznego, który jest kompatybilny z definicją MDC.



Rycina 6. Schemat wyjaśniania mechanistycznego⁴⁹

Górna część schematu reprezentuje to, co ma zostać wyjaśnione (*explanandum*)⁵⁰. Symbol ψ (od *psychological*) oznacza zjawisko, własność bądź zachowanie, wyjaśniane przez mechanizm. Z kolei symbol S oznacza mechanizm jako całość. Dolna część schematu reprezentuje obiekty (okręgi) oraz działania (strzałki wychodzące od okręgów). Zgodnie z definicją MDC, wspólnie konstytuują one mechanizm. W wyjaśnianiu pełnią one rolę *explanansu* (tego, co wyjaśnia).

⁴⁸ Zob. A. Revonsuo, *O naturze wyjaśniania w neuronaukach*, op. cit., s. 279.

⁴⁹ Rycina sporządzona na podstawie: C.F. Craver, *Explaining the Brain...*, op. cit., s. 7.

⁵⁰ *Ibidem*, s. 6–7.

Symbol X oznacza komponent (część) obiektu, zaś ϕ (od *physiological*) działanie danego obiektu w mechanizmie. Wyjaśnienie zjawiska $S\psi$ polega na odkryciu i pokazaniu struktury obiektów $\{X_1, X_2, X_3, \dots, X_m\}$ oraz działań $\{\phi_1, \phi_2, \phi_3, \dots, \phi_n\}$.

Wróćmy teraz do wspomnianego wyżej mechanizmu neuroprzekąźnikowego. Impuls nerwowy jest przekazywany na drodze elektrycznej oraz chemicznej. Jak wiadomo, sygnał elektryczny z komórki presynaptycznej jest zamieniany na sygnał chemiczny, a następnie w komórce postsynaptycznej znów na sygnał elektryczny. Praktycznie w każdym podręczniku do neuronauki czy biologicznych podstaw psychologii mechanizm neuroprzekąźnikowy jest reprezentowany jako dwuwymiarowy *diagram* (oczywiście mniej lub bardziej nasycony szczegółami). Diagramy te przeważnie można zinterpretować – zgodnie z definicją MDC – jako *obiekty* wraz z *własnościami* oraz *działaniami*. Obiektami są między innymi błony komórkowe, membrany, mikrotubule, molekuly i jony. Działania są natomiast reprezentowane za pomocą strzałek, przerywanych linii itd. Odnoszą się one do zjawisk takich jak biosynteza, depolaryzacja, dyfuzja czy też modulacja.

Machamer, Darden i Craver szczegółowo omawiają depolaryzację błony komórkowej⁵¹. Ich zdaniem przykład ten pozwala przybliżyć typowe cechy wyjaśniania mechanistycznego⁵². Jak wiadomo, błona komórki nerwowej zachowuje stan polaryzacji. Oznacza to, że wewnątrz komórki posiada ujemny ładunek w stosunku do jej zewnątrz. Różnica napięcia elektrycznego z obydwu stron błony nieaktywnego neuronu jest określana jako potencjał spoczynkowy. Potencjały spoczynkowe neuronów wynoszą w dużym przybliżeniu -70 mV (w istocie przybierają wartości z przedziału od -65 mV do -90 mV). Depolaryzacja polega na zmniejszeniu elektroujemnego potencjału

⁵¹ Szczegółowe omówienie procesu depolaryzacji można znaleźć w praktycznie dowolnym podręczniku do neuronauki lub biopsychologii. Zob. A. Longstaff, *Neurobiologia...*, *op. cit.*; J. Kalat, *Biologiczne podstawy psychologii*, *op. cit.*, s. 39–44.

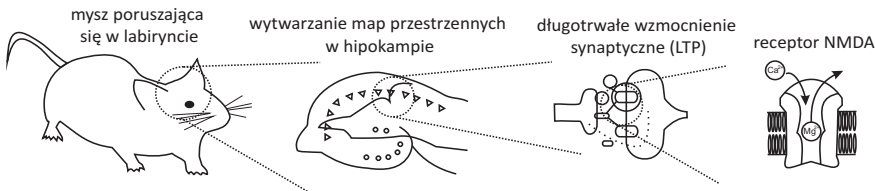
⁵² P.K. Machamer, L. Darden, C.F. Craver, *Thinking about Mechanisms*, *op. cit.*, s. 8–11.

błony. Zjawisko to jest związane ze zmianami w dystrybucji jonów. Błona komórkowa jest selektywnie przepuszczalna. W zależności od warunków kanały mogą przepuszczać jony lub nie. Przepływ jonów reguluje pompa sodowo-potasowa, która wydalą na zewnątrz neuronu trzy jony Na^+ , wpuszczając do środka dwa jony K^+ . Napływ jonów Na^+ do wnętrza komórki prowadzi do zwiększenia potencjału wewnątrz neuronu. Jeśli komórka (o potencjale spoczynkowym) zostanie pobudzona, powstaje potencjał czynnościowy (informacja jest przykazywana do neuronu postsynaptycznego). Mechanizm ten, podobnie jak każdy inny, jest charakteryzowany przez warunki początkowe, pośrednie oraz końcowe.

Warunki początkowe (*set-up conditions*) mogą być rezultatem wcześniejszych procesów, jednak zazwyczaj – na mocy idealizacji – są one utożsamiane z początkiem danego procesu⁵³. Na warunki początkowe składa się odpowiednia konfiguracja obiektów i ich własności (obejmuje ona np. relacje przestrzenne między obiektami). Procesy są uruchamiane dzięki warunkom aktywacji (*enabling conditions*), których rolę może pełnić konfiguracja i dystrybucja jonów Na^+ , a także czynniki takie jak temperatura oraz pH. Z kolei warunki końcowe (*termination conditions*) opisują końcową fazę, czy też wygaśnięcie procesu. W rezultacie pewne obiekty mogą być eliminowane, mogą pojawiać się nowe obiekty, układ może osiągnąć stan równowagi itd. Oczywiście jest to idealizacja, gdyż koniec pewnego procesu jest zwykle początkiem nowych procesów biologicznych. W przypadku mechanizmu depolaryzacji warunki te odnoszą się do wzrostu poziomu jonów Na^+ wewnątrz neuronu. Koreluje to ze wzrostem napięcia elektrycznego błony komórkowej. Pomiędzy początkiem a końcem tego procesu znajduje się jednak wiele, zachodzących w sposób ciągły, procesów pośredniczących (*intermediate activities*). W przypadku depolaryzacji wyjaśnienie dostarczyć musi między innymi dokładnego opisu wnikania jonów Na^+ przez kanały.

⁵³ Zob. *ibidem*, s. 11.

Rozważmy pobieżnie jeszcze jeden przykład – zaproponowany przez Cravera i Darden – który odnosi się do wyjaśnienia mechanizmu pamięci przestrzennej myszy⁵⁴. Liczne badania wskazują, że mapy przestrzenne są zakodowane w hipokampie, należącym do układu limbicznego. Ich tworzenie jest możliwe dzięki długotrwałym wzmocnieniom synaptycznym (*long-term potentiation*, LTP), w których istotną rolę odgrywa z kolei aktywacja receptorów NMDA (główną funkcją tych receptorów jest regulacja dystrybucji Ca^+ do wnętrza neuronu). Przykład ten pokazuje, że wyjaśnianie mechanizmów jest *poziomowe*. Oznacza to, że wyjaśnienie określonego zjawiska polega na odkryciu i przedstawieniu całej *hierarchii*, poczynawszy od wyjaśnianego zjawiska, a skończywszy na najniższym poziomie, na którym znajdują się obiekty i procesy konstytuujące dany mechanizm. W przypadku mechanizmu pamięci przestrzennej aktywacja receptorów NMDA jest procesem niższego poziomu dla LTP, z kolei LTP jest procesem niższego poziomu dla konstruowania map przestrzennych w hipokampie, zaś to jest mechanizmem niższego poziomu dla pamięci przestrzennej myszy⁵⁵. Hierarchię omawianego mechanizmu ilustruje poniższa rycina.



Rycina 7. Poziomy hierarchicznej organizacji mechanizmu pamięci przestrzennej myszy⁵⁶

⁵⁴ C.F. Craver, L. Darden, *Discovering Mechanisms in Neurobiology. The Case of Spatial Memory*, [w:] *Theory and Method in the Neurosciences*, red. P.K. Machamer, R. Grush, P. McLaughlin, University of Pittsburgh Press, Pittsburgh 2001, s. 112–137; zob. także C.F. Craver, *The Making of a Memory Mechanism*, „Journal of the History of Biology” 2003, vol. 36, no. 1, s. 153–195.

⁵⁵ Zob. C.F. Craver, L. Darden, *Discovering Mechanisms in Neurobiology...*, *op. cit.*, s. 117.

⁵⁶ Rycina została przygotowana na podstawie: *ibidem*, s. 118.

Przykład ten pokazuje, że wyjaśnianie mechanistyczne może operować na wielu poziomach teoretycznych. Nie jest ono redukcjonistyczne, a raczej zmierza do stworzenia sieci łączącej różne poziomy złożoności. Craver pisze zresztą wprost, że wyjaśnienia neuronaukowe mają charakter wielopoziomowy. Przydatną dystynkcję wprowadza José Luis Bermúdez. Opis pojedynczego poziomu określa on mianem *wyjaśnienia horyzontalnego*, zaś przejścia między różnymi poziomami w hierarchii teoretycznej nazywa *wyjaśnieniem wertykalnym*⁵⁷.

Powyżej przedstawiłem ogólny zarys wyjaśniania mechanistycznego, które zalicza się do semantycznych strategii wyjaśniania naukowego. Omówiłem także podstawowe pojęcia z mechanistycznego słownika oraz zobowiązania ontologiczne mechanicyzmu. Większość podawanych w literaturze przykładów wyjaśniania mechanistycznego odnosi się do poziomu pojedynczych komórek oraz bogatszych struktur komórkowych. Z drugiej strony, jak pisze Antti Revonsuo: „w neuronauce kognitywnej najwyraźniej próbuje się rozszerzyć zakres zastosowania strategii biologicznego wyjaśniania mechanistycznego aż na zjawiska psychologiczne”⁵⁸. Udany przykładem takiej próby jest opisany powyżej mechanizm pamięci przestrzennej. Rozszerzenie mechanicyzmu na cały poziom psychologiczny, czy też (zgodnie z terminologią przyjętą w niniejszej pracy) na poziom procesów poznawczych, napotyka jednak na problemy.

Przeciw wyjaśnianiu mechanistycznemu w neuronauce formułuje się szereg zarzutów. Niektóre z nich odnoszą się do niższych poziomów opisu, czyli pojedynczych neuronów oraz bogatszych struktur neuronalnych, inne zaś są specyficzne dla wyjaśniania procesów poznawczych. Jeśli chodzi o niższe poziomy analizy neurokognitywnej, Revonsuo wskazuje zarówno trudności, które dotyczą istoty mechanicyzmu, jak i różnego rodzaju problemy epistemologiczne⁵⁹.

⁵⁷ Zob. J.L. Bermúdez, *Philosophy of Psychology*, op. cit., s. 32–33.

⁵⁸ A. Revonsuo, *O naturze wyjaśniania w neuronaukach*, op. cit., s. 287.

⁵⁹ Zob. *ibidem*, s. 278–279.

Pierwsza trudność jest związana z procedurą dekompozycji, czyli wyodrębniania elementów mechanizmu i rozpoznawania pełnionych przez nie funkcji. Trudność ta pojawia się na poziomie pojedynczych komórek, które – jak wiadomo – są strukturami „nieliniowymi”, tj. takimi, których całość jest czymś więcej niż sumą pojedynczych elementów. Problem staje się jeszcze bardziej doniosły, gdy próbujemy poddać dekompozycji bogatsze struktury neuronalne. Jak pisze Revonsuo:

Jedynie systemy złożone z modularnych podsystemów są w tym sensie rozkładalne – gdy idzie o własności podsystemów, przyczynowe interakcje *w obrębie* podsystemu muszą być bardziej istotne niż interakcje *pomiędzy* podsystemami⁶⁰.

Wyjaśnianie mechanistyczne jest w naturalny sposób związane z tezą o modularności mózgu i umysłu. W pierwszym rozdziale niniejszej pracy pokazałem, że modularność jest uwikłana w szereg problemów. O ile pojedyncze komórki nerwowe oraz struktury podkorowe można uważać za poddające się dekompozycji, problem pojawia się na poziomie struktur korowych. Z analizy przeprowadzonej w pierwszym rozdziale wynika, że trudno jest zgodzić się, iż organizacja mózgu, przynajmniej na poziomie kory mózgowej, ma charakter modularny w sensie anatomicznym. Z drugiej strony, mechanistyczna strategia wyjaśniania jest akceptowalna dzięki heurystyce modularności. O ile w ogóle można mówić o modularności na poziomie kory, o tyle znacznie logiczniejszym ujęciem jest modularność w sensie sieciowym. Znow jednak mamy do czynienia z przeszkodą, tyle że poznawczą. Dekompozycja modułów w sensie sieciowym jest bardzo trudna to przeprowadzenia. Jeśli chodzi o system neuronalny, Revonsuo konkluduje:

Pojawiły się (...) trudności z empiryczną identyfikacją przyczynowo istotnych części tego systemu i problemy te mogą świadczyć o tym, że

⁶⁰ *Ibidem*, s. 279.

mózg nie daje się rozłożyć na każdym poziomie organizacji w stopniu wystarczającym dla podania mechanistycznych wyjaśnień⁶¹.

Nawet jeśli dekompozycja jest w ogóle możliwa, jest ona bardzo silnie uwikłana w przyjmowaną przez danego badacza teorię. W tym kontekście Revonsuo wymienia choćby różne teorie przetwarzania informacji przez mózg. Najogólniej mówiąc, teoria przetwarzania neuronalnego powinna opisać mechanizm, którego działanie pozwala na transmisję informacji sensorycznej, jej transformację w informację percepcyjną i/lub motoryczną oraz wygenerowanie zachowania organizmu⁶². Teoria ma więc wyjaśniać drogę od „pobrania” danych ze środowiska aż do obserwowalnej intersubiektywnie reakcji organizmu. Wciąż nie ma jednak zgody, jak to się dzieje. Dobrze ilustrują to słowa Patricii Churchland:

Nie rozumiemy, co pojęcie *informacji* powinno oznaczać w kontekście biologicznym czy psychologicznym. Co więcej, nie rozumiemy jeszcze w pełni, jak neurony kodują informacje, czymkolwiek informacja *jest*⁶³.

Revonsuo referuje trzy konkurencyjne teorie przetwarzania informacji. Pierwsza z nich to teoria komórek babcinych (*grandmother cell theory*)⁶⁴. Mówi ona, że każda jednostka percepcyjna jest kodowana przez jedną konkretną komórkę nerwową. Komórki są zorganizowane hierarchicznie. Neurony odpowiedzialne za przetwarzanie prostych informacji przesyłają rezultaty swojej pracy do wyższych hierarchicznie komórek. W wyniku przetwarzania informacji przez całą hierarchię neuronów są generowane obiekty będące przedmiotem percepcji.

⁶¹ *Ibidem*.

⁶² Zob. *ibidem*, s. 281.

⁶³ P.S. Churchland, *Brain-Wise. Studies in Neurophilosophy*, The MIT Press, Cambridge–London 2002, s. 170.

⁶⁴ Zob. A. Revonsuo, *O naturze wyjaśniania w neuronaukach*, *op. cit.*, s. 281.

Innego rozwiązania dostarcza teoria zespołów komórkowych (*cell assembly theory*), która mówi o sieciowym kodowaniu obiektów percepcyjnych. Każdy taki obiekt jest reprezentowany jako wzorzec aktywności sieci. W odróżnieniu od teorii komórek babcinych, zwolennicy ujęcia sieciowego uważają, że pojedyncze neurony nie kodują jednostek percepcyjnych, ale elementarne *własności*. Kluczową rolę w tworzeniu reprezentacji odgrywają zmiany wartości pobudzeń komórek nerwowych. Różne konfiguracje własności mogą reprezentować różne obiekty, stąd też jeden zespół komórkowy może uczestniczyć w percepcji bardzo dużej liczby różnych obiektów.

Z kolei teoria kodowania czasowego (*temporal coding theory*) mówi, że percepcja nie może być wyjaśniona jedynie za pomocą zmian pobudzeń neuronów. Dwie poprzednie teorie mówiły o kodowaniu jednostek percepcyjnych. W horyzoncie percepcyjnym niemal zawsze pojawia się jednak więcej niż jedna reprezentacja obiektu. Zdaniem zwolenników teorii kodowania czasowego, neurony należące do danego wzoru są aktywowane symultanicznie, tworząc reprezentację pojedynczego obiektu. W tym samym czasie aktywne mogą być również neurony należące do innych wzorów, generujące inne reprezentacje. Revonsuo pisze, że „(...) ich konstytutywne komórki muszą wyznaczać odmienny rytm synchronicznego pobudzenia”⁶⁵.

Teoria komórek babcinych, poddana krytyce przez Jerome’a Lettina, ma już niemal jedynie znaczenie historyczne. Bardziej spójne z wynikami różnych eksperymentów wydają się dwie pozostałe koncepcje, jednak żadna z nich nie jest traktowana jako ostateczna teoria przetwarzania informacji. Podstawowe trudności są związane z oddzieleniem procesów relewantnych przyczynowo na danym poziomie od procesów, które można zaniedbać. Oczywiście jest to kolejna odsłona problemu idealizacji w neuronauce poznawczej. Dotykamy tu kwestii związku między poziomem pojedynczych neuronów oraz bogatszych struktur neuronalnych. Revonsuo przedstawia

⁶⁵ *Ibidem*; zob. także M.S. Gazzaniga, R.B. Ivry, G.R. Mangun, *Cognitive Neuroscience...*, *op. cit.*

w tym kontekście dwa wyróżniane przez Theodore'a Holmesa Bullocka poglądy na temat komunikacji struktur mózgowych: *perspektywę jednostkową (unit window view)* oraz *perspektywę populacyjną (population window view)*. W perspektywie jednostkowej komunikację niemal wszystkich struktur neuronalnych próbuje się wyjaśnić redukcyjnie w odniesieniu do pojedynczych neuronów. Również komunikację wyższych struktur neuronalnych tłumaczy się adekwatnością „próbkowania sygnału pochodzącego z neuronu”⁶⁶. Z kolei w ujęciu populacyjnym komunikacji całego systemu nerwowego nie można zredukować do aktywności pojedynczych neuronów. Na poziomie bogatszych struktur neuronalnych pojawiają się bowiem *nowe własności przyczynowe*. Chodzi tu o wolno zmieniające się potencjały, poziomy i rytmy synchronizacji oraz całościowe wzorce aktywacji tych struktur. Druga z perspektyw wydaje się bardziej atrakcyjna, jednak stwarza spore trudności dla mechanistycznej strategii eksplanacyjnej. Z racji ogromnej komplikacji sieci neuronalnej trudno jest oddzielić zjawiska relewantne funkcjonalnie od zjawisk zaniedbywalnych. Revonsuo wyciąga następujący wniosek:

Problem przetwarzania informacji neuronowej jest zatem przejawem bardziej fundamentalnego problemu związanego z mechanistyczną strategią eksplanacyjną: problemu identyfikacji funkcjonalnie istotnych części systemu nerwowego i poziomów organizacji przez nie tworzonych (...). Trudno rozstrzygnąć, czy jest to jedynie wynik obecnego stanu wiedzy czy też znak, że mózg może nie spełniać założeń mechanistycznego modelu wyjaśniania⁶⁷.

Inny przykład w identyfikacji struktur istotnych przyczynowo jest związany z pytaniem o ilość autonomicznych obszarów kory wzro-

⁶⁶ Cyt. za: A. Revonsuo, *O naturze wyjaśniania w neuronaukach*, op. cit., s. 282 (T.H. Bullock, *Signals and Signs in the Nervous System. The Dynamic Anatomy of Electrical Activity Is Probably Information-Rich*, „Proceedings of the National Academy of Sciences of the U.S.A.” 1997, no. 94, s. 1–6).

⁶⁷ *Ibidem*, s. 283.

kowej⁶⁸. Problem ten wydaje się o tyle poważny, że układ wzrokowy jest prawdopodobnie najlepiej zbadanym „składnikiem” systemu poznawczego człowieka. O ile panuje zgoda co do *lokalizacji* oraz *funkcji* obszarów V1 (czyli pierwszorzędowej kory wzrokowej) oraz V2, V3/VP i MT, o tyle nie ma już takiej zgody co do ilości oraz funkcji wyższych obszarów kory wzrokowej, takich jak V4, V8 czy też V4V. Wyróżnienie któregoś obszaru jako (względnie) autonomicznego jest niezwykle trudne ze względu na dużą ilość połączeń z obszarami zlokalizowanymi poza korą prążkową. Silnie rozwinięte połączenia pomiędzy różnymi obszarami sprawiają, że trudno jest w nich wyodrębnić samodzielne anatomicznie i funkcjonalnie struktury, które można zdekomponować. Nie wiadomo ponadto, jakich kryteriów należałoby użyć, aby uznać, że dany obszar wyższej kory wzrokowej jest samodzielny. Standardowe kryteria używane do identyfikacji niższych obszarów kory, takie jak kryteria retinotopowe, histologiczne oraz koneksjonistyczne, nie zdają egzaminu w wypadku wyższych obszarów. Oczywiście, zaprezentowane w pierwszym rozdziale niniejszej pracy argumenty podające w wątpliwość silną – anatomiczną – wersję modularności mózgu stanowią również poważne wyzwanie dla mechanistycznej strategii wyjaśniania.

Do mechanistycznej strategii wyjaśniania, opartej na definicji MDC, odnoszą się ponadto wprost Barbara Von Eckardt oraz Jeffrey S. Poland⁶⁹. Ich zdaniem, choć mechanicyzm w wielu przypadkach jest dobrą strategią wyjaśniania w neuronauce poznawczej, nie jest strategią omnipotentną. Mechanicyzm „w szczególności nie jest w stanie uchwycić w całościowy sposób wszystkich aspektów treści i znaczenia reprezentacji mentalnych oraz istotnych cech klasyfikacyjnych w neuropatologii”⁷⁰. Zdaniem Von Eckardt i Polanda, pojęcie *reprezentacji* pełni bardzo ważną rolę w wyjaśnieniach dostarczanych przez neuronaukę poznawczą. Nauka ta jest zainteresowana

⁶⁸ Zob. *ibidem*, s. 284.

⁶⁹ Zob. B. Von Eckardt, J.S. Poland, *Mechanism and Explanation in Cognitive Neuroscience*, „Philosophy of Science” 2004, vol. 71, no. 5, s. 972–984.

⁷⁰ *Ibidem*, s. 972.

wyjaśnianiem zjawisk z poziomu psychologicznego, czy też poziomu procesów poznawczych. Ich zdaniem poziom ten obejmuje *stany intencjonalne*. Przyjmijmy za Johnem Searle'em, że: „'intencjonalność' oznacza skierowanie umysłu na coś (*directedness*), czyli to, że każdy stan umysłu czegoś dotyczy (*aboutness*)”⁷¹. Zdaniem Von Eckardt i Polanda, kognywiści wyjaśniają intencjonalność, stosując strategię 'semantycznej dziedziczności' (*semantic heritability*):

Strategia ta wyjaśnia intencjonalność przez odniesienie do reprezentacji, które niosą treść i znaczenie dla ich „właściciela”. Zgodnie z tą strategią wyjaśnienie, jak osoba wykazuje pewną zdolność (np. czyta zdanie), odnosi się do historii komputacyjnej ('symbolicznej' lub konksjonistycznej), ukazującej kroki czy też etapy, jakie przebywa umysł/mózg tej osoby od reprezentacji wzrokowej (lub słuchowej) napisu lub wymowy tego zdania do reprezentacji jego znaczenia, to znaczy stanu będącego podłożem dla naszego „rozumienia” zdania. Wyjaśnienie, w jaki sposób osoba *posiada* (...) pewną zdolność, odnosi się do obliczeniowych i reprezentacyjnych zasobów, niezbędnych dla korzystania z tej zdolności⁷².

Von Eckardt i Poland uważają, że uprawianie neuronauki poznawczej polega na połączeniu kognitywistyki *sensu stricto* i „podstawowej” neuronauki. Teorie neurokognitywne są wówczas konstruowane poprzez „rzutowanie” obliczeniowych modeli umysłu na struktury neuronalne. Pojęcie reprezentacji wykorzystywane w wyjaśnianiu kognitywistycznym zaczyna wówczas dotyczyć w automatyczny sposób całej neuronauki poznawczej. Von Eckardt i Poland argumentują, że pytanie o *realizację* treści i znaczenia prowadzi do ujawnienia niespójności pomiędzy podejściem mechanistycznym i reprezentacyjnym⁷³. Najogólniej chodzi o to, że reprezentacje mają

⁷¹ J.R. Searle, *Umysł. Krótkie wprowadzenie*, *op. cit.*, s. 14; zob. także rozdz. 7.

⁷² Zob. B. Von Eckardt, J.S. Poland, *Mechanism and Explanation in Cognitive Neuroscience*, *op. cit.*, s. 979.

⁷³ Zob. *ibidem*, s. 981–982.

komponent eksternalny. Ich zdaniem mechanistyczne wyjaśnienie wszystkich zdolności kognitywnych nie jest możliwe, gdyż „reprezentacje nie są w pełni tworzone (realizowane) na niższym poziomie neuronalnych obiektów i procesów (...)”⁷⁴. Innymi słowy, mechanicyzm nie jest w stanie uchwycić wszystkich treściowych i znaczeniowych aspektów reprezentacji.

Szczególny przypadek omówionej wyżej trudności mechanicyzmu odnosi się do wyjaśniania w psychopatologii (neuropatologii). Von Eckardt i Poland zwracają uwagę, że w psychiatrii przez *urojenia (delusions)* rozumie się uporczywe utrzymywanie *falszywych przekonań* na temat *zewnętrznej rzeczywistości*⁷⁵. Urojenia są zatem rozumiane w sposób intencjonalny. Wyjaśnianie przyczyn wystąpienia u danej osoby urojeń odnosi się do treści przekonań, a więc ma charakter reprezentacyjny. Prowadzi to do powyższego zarzutu: jeśli wyjaśnienia neuropatologiczne mają charakter intencjonalny, w sposób mechanistyczny nie da się uchwycić wszystkich aspektów treści i znaczenia urojeń. To jednak nie koniec problemów. Von Eckardt i Poland podkreślają, że klasyfikacja zjawisk interesujących psychiatrów ma charakter normatywny. Oznacza to, że aby stwierdzić, czy (a jeśli tak, to dlaczego) dana osoba cierpi na urojenia, trzeba porównać jej zdolności poznawcze oraz zachowanie z pewną normą. Norma ta może być stanowiona przez środowisko psychiatrów, lub choćby wywodzić się z potocznej obserwacji zachowań zdrowych ludzi. Jeśli wyjaśnianie neuropatologiczne rzeczywiście odwołuje się do norm oraz porównywania różnych przypadków, staje się problematyczne dla mechanicyzmu.

Uogólniając zarzuty Von Eckardt i Polanda: mechanicyzm może radzić sobie bardzo dobrze na poziomie „podstawowej” neuronauki, ale zawodzi na poziomie specyficznym dla neuronauki poznawczej. Moim zdaniem ich zarzuty można jednak osłabić. Po pierwsze, twierdzenie, że wyjaśnianie ma charakter reprezentacyjny nie odnosi się

⁷⁴ *Ibidem*, s. 983.

⁷⁵ *Zob. ibidem*, s. 982.

do całej neuronauki poznawczej, ale jedynie do paradygmatu komputerowego. Związane jest z tym moje drugie zastrzeżenie: neuro nauka poznawcza jest przez nich rozumiana w niezbyt wyrafinowany sposób, jako swoista hybryda kognitywistyki *sensu stricto* i „podstawowej” neuronauki. Wreszcie po trzecie, zgodnie z przyjętą przeze mnie ramą pojęciową, psychiatria nie należy do neuronauki poznawczej *prima facie*, ale zalicza się do jej wiedzy towarzyszącej. Niezależnie od moich obiekcji zgadzam się z Von Eckardt i Polandem co do tego, że mechanicyzm nie obejmuje wszystkich aspektów teorii neurokognitywnych. Uważam ponadto, że nie jest on fundamentalną strategią wyjaśniania. Aby to pokazać, przyjrzyjmy się teraz pewnym podobieństwom mechanicyzmu oraz innych strategii wyjaśniania.

Na pierwszy rzut oka największą różnicą w wyjaśnianiu dedukcyjno-nomologicznym i mechanistycznym są *explanansy*. Jak wiadomo, w przypadku wyjaśniania w duchu neopozytywistycznym w *explanansie* musi się znajdować co najmniej jedno prawo. Paradoxy wyjaśniania uzmysłowiły ponadto, że w wyjaśnianiu nie można zaniedbywać zdań odnoszących się do związków przyczynowych. Z kolei w przypadku mechanicyzmu wyjaśnianie opiera się na dekompozycji, która pozwala dotrzeć do obiektów i działań. Nie wszyscy zwolennicy mechanicyzmu sprzeciwiają się jednak wykorzystywaniu *praw* w wyjaśnianiu. Przykładowo Stuart S. Glennan pisze, że:

Mechanizm będący podłożem zachowania jest złożonym systemem, który generuje to zachowanie (...) poprzez interakcje wielu części, rządzone bezpośrednimi prawami przyczynowymi⁷⁶.

Do zagadnienia praw powrócę jeszcze w końcowej części tego rozdziału. Prawie wszyscy zwolennicy mechanicyzmu mają na uwadze *przyczynowość*. Pojawia się ona w koncepcji Machamera, Darden i Cravera. Ich zdaniem „działania (*activities*) są rodzajami

⁷⁶ S.S. Glennan, *Mechanisms and the Nature of Causation*, „Erkenntnis” 1996, no. 44, s. 52.

przyczyn⁷⁷. W innej pracy Craver pisze z kolei, że „działania są przyczynowymi komponentami mechanizmów⁷⁸. Autorzy ci przywołują poglądy Gertrude Elizabeth Margaret Anscombe, która twierdzi, że pojęcie *przyczyny* jest bardzo ogólne i staje się zrozumiałe dopiero po dodaniu odpowiedniego czasownika, wyrażającego działanie przyczynowe (np. *popycha, przyciąga, uderza* itd.)⁷⁹. Co więcej, Machamer, Darden i Craver twierdzą, że poglądy Johna Lesliego Mackiego na naturę przyczynowości mogą być zinterpretowane w duchu mechanistycznym⁸⁰.

Pozostaje omówić jeszcze jedną różnicę, która dzieli koncepcje syntaktyczne i semantyczne. W pierwszych wyjaśnianie polega na przeprowadzeniu operacji na *zdaniach*, zaś w drugich na odkryciu i pokazaniu działania mechanizmu. Bardzo często, jak twierdzi William Bechtel, lepsze wyjaśnienie interesującego zjawiska jest uzyskiwane, gdy mechanizm zostanie zaprezentowany w formie *graficznej* (wykresy, diagramy, schematy)⁸¹. Sądzę jednak, że różnica ta jest pozorna. Analizując mechanizm neurotransmisji, Machamer, Darden i Craver przyznają: „(...) diagramy reprezentują cechy mechanizmów, które mogą być opisywane werbalnie, ale znacznie łatwiej ująć je w formie wizualnej⁸². Dobrym przykładem jest opisana przeze mnie w rozdziale drugim architektura kognitywna ACT*. Działanie tej architektury najłatwiej wyjaśnić w sposób graficzny, ale nie ma żadnej przeszkody, by w pełni opisać je zdaniami. Preferencja postaci graficznej odnosi się więc jedynie do ekonomii poznawczej. Zastosowanie modeli ikonicznych nie prowadzi w konieczny sposób do odrzucenia zdaniowej koncepcji teorii neurokognitywnych i przyjęcia

⁷⁷ P.K. Machamer, L. Darden, C.F. Craver, *Thinking about Mechanisms*, *op. cit.*, s. 6.

⁷⁸ C.F. Craver, *Explaining the Brain...*, *op. cit.*, s. 6.

⁷⁹ Zob. P.K. Machamer, L. Darden, C.F. Craver, *Thinking about Mechanisms*, *op. cit.*, s. 6; G.E.M. Anscombe, *Causality and Determination. An Inaugural Lecture*, Cambridge University Press, Cambridge 1971.

⁸⁰ Zob. *ibidem*, s. 7; J.L. Mackie, *The Cement of the Universe. A Study of Causation*, Oxford University Press, Oxford 1974, *passim* (zwl. s. 213–229).

⁸¹ Zob. W. Bechtel, R.C. Richardson, *Discovering Complexity...*, *op. cit.*

⁸² P.K. Machamer, L. Darden, C.F. Craver, *Thinking about Mechanisms*, *op. cit.*, s. 8.

niezdaniowego ujęcia teorii. Co więcej, łatwo znaleźć przykłady problemów, gdzie wyjaśnienie zdaniowe jest znacznie bardziej ekonomiczne niż tworzenie reprezentacji graficznej. Przykładowo, wątpliwość w sensowność tworzenia graficznej reprezentacji działania modułu wykrywania oszustów (modułu normatywnego), którego istnienie postulują psychologowie ewolucyjni.

Rozważmy teraz niezwykle interesujący przykład z historii nauki, który świadczy o tym, że różnego rodzaju reprezentacje graficzne co do zasady mogą być przełożone na zdania w sposób nietrywialny. Rzecz dotyczy dowodów twierdzeń w *Elementach* Euklidesa. Ponieważ przykład ten odnosi się do matematyki, a nie nauk biologicznych, można go potraktować jako dygresję kończącą niniejszy podrozdział. Zaprezentuję go w dużym uproszczeniu, bez wchodzenia w techniczne szczegóły. System aksjomatyczno-dedukcyjny skonstruowany i opisany przez Euklidesa w *Elementach* uchodzi powszechnie za wzór ścisłości. Do systemu Euklidesa odwoływali się później niemal wszyscy filozofowie, którzy chcieli udowodnić coś *more geometrico*. Dzieło Euklidesa można uznać za pionierskie dla sformułowanej przez Hilberta idei systemu formalnego (pojęcie to omówiłem pobieżnie w pierwszym rozdziale). Z racji tego, że wielu filozofów nauki zgłaszało roszczenia, by teorie naukowe przypominały systemy aksjomatyczno-dedukcyjne (tyle że wzbogacone o treść empiryczną), *Elementy* można również potraktować jako archetyp teorii naukowej w ogóle. Grecki uczyony usystematyzował wiedzę matematyczną między innymi w dziedzinach geometrii i teorii liczb. Jak wiadomo, Euklides ufundował geometrię na pięciu postulatach (powszechnie, ale mylnie, nazywane są dziś one aksjomatami):

- (1) Od każdego punktu do każdego punktu można poprowadzić prostą;
- (2) Ograniczoną prostą można w sposób ciągle przedłużać wzdłuż prostej;
- (3) Z każdego środka każdą rozwartością można opisać koło;
- (4) Wszystkie kąty proste są sobie równe;

- (5) Jeśli prosta, przecinająca dwie proste, tworzy kąty wewnętrzne jednostronnie mniejsze od dwóch prostych, to te dwie proste, przedłużane nieograniczenie, spotkają się z tej strony, gdzie kąty są mniejsze od dwóch prostych⁸³.

Teorią przedstawioną w *Elementach* wstrząsnęły (co najmniej) dwa zdarzenia. Po pierwsze, okazało się, że geometria euklidesowa nie jest geometrią absolutną (jak twierdził np. Immanuel Kant). Problematyczny okazał się przede wszystkim piąty postulat. Próby wyprowadzenia go z pozostałych czterech nie zakończyły się sukcesem. Poprzez negację piątego postulatu otrzymano całą gamę geometrii nieeuklidesowych. Znalazły one wiele zastosowań, szczególnie w kosmologii. Dla problematyki wyjaśniania naukowego znacznie istotniejsze jest jednak coś innego. Aksjomatyka Euklidesa okazała się *niepełna*. W roku 1882 Moritz Pasch znalazł przykład prawdziwej formuły, której nie można udowodnić na gruncie aksjomatów sformułowanych przez Euklidesa. Pasch pokazał, że aksjomatyczne ujęcie geometrii euklidesowej zawiera dodatkowy, nieujawniony aksjomat (nazywany dziś *aksjomatem Pascha*): „prosta nieprzechodząca przez żaden z wierzchołków trójkąta, ale przecinająca jeden bok trójkąta, musi przeciąć jeszcze przynajmniej jeden bok”⁸⁴. Okazało się, że wiele rozumowań przedstawionych w *Elementach* ma charakter entymematyczny. Oznacza to, że Euklides wykorzystywał entymematy, czyli nieujawnione przesłanki. W większości przypadków głównym motywem było uznanie pewnych zdań za *oczywiste*⁸⁵. Istotną rolę w procesie dowodowym pełniły również własności *diagramów* i *rysunków pomocniczych*. „Graficzne” podejście do geometrii jest dobrze znane wszystkim z edukacji szkolnej, pozostaje jednak dalekie od ścisłości. Odkrycie Pascha stymulowało do wypracowania w pełni syntaktycznego ujęcia geometrii:

⁸³ J. Dadaczyński, *Filozofia matematyki w ujęciu historycznym*, Biblos, Tarnów 2000, s. 94.

⁸⁴ K. Wójtowicz, *O pojęciu dowodu w matematyce*, op. cit., s. 31.

⁸⁵ Zob. R. Murawski, *Filozofia matematyki. Zarys dziejów*, Wyd. Naukowe PWN, Warszawa 2001, s. 204–216 (Dodatek II: *Kilka uwag o filozofii geometrii*).

Jeśli geometria ma być nauką dedukcyjną, proces wnioskowania musi we wszystkich fragmentach być niezależny od znaczenia pojęć geometrycznych, podobnie jak musi być niezależny od diagramów; pod uwagę mogą być brane jedynie relacje wyrażane w twierdzeniach i definicjach. W czasie wnioskowania jest użyteczne i dopuszczalne, ale nie konieczne myślenie o znaczeniach terminów; faktycznie, jeśli jest to konieczne, to w ten sposób widoczna staje się niepoprawność dowodu⁸⁶.

Dla jasności należy podkreślić, że rozumowania przeprowadzone przez Euklidesa były poprawne, tyle że niezgodne z ideą systemu aksjomatyczno-dedukcyjnego. Innymi słowy, uwzględniały one elementy semantyczne, podczas gdy od drugiej połowy XIX wieku zaczęła dominować tendencja do syntaktycznego (formalnego) traktowania geometrii⁸⁷. Dał temu wyraz David Hilbert w swoich słynnych słowach: „Zawsze musi być możliwe podstawienie ‘stołów’, ‘krzesel’ czy ‘kufli’ za ‘punkty’, ‘proste’ czy ‘płaszczyzny’ w systemie aksjomatów Euklidesa”⁸⁸. Dla uzyskania pełności aksjomatykę podaną przez Euklidesa w *Elementach* trzeba było uzupełnić między innymi o postulaty stereometryczne, aksjomaty ruchu i aksjomaty ciągłości⁸⁹. W 1899 roku Hilbert wykazał w słynnej pracy *Grundlagen der Geometrie*, że geometria euklidesowa jest oparta nie na pięciu, ale ponad dwudziestu aksjomatach. Dla interesującego nas zagadnienia oznacza to, że Hilbertowi udało się „przetłumaczyć” zawierający elementy semantyczne (*implicite*) system Euklidesa na system czysto syntaktyczny. Wszystkie miejsca, gdzie były wykorzystywane elementy graficzne, zostały przełożone w nietrywialny sposób na zdania. Osobnym zagadnieniem pozostaje, czy po tej operacji geometria

⁸⁶ Cyt. za: K. Wójtowicz, *O pojęciu dowodu w matematyce*, op. cit., s. 31 (M. Pasch, *Vorlesungen über neuere Geometrie*, B.G. Teubner, Leipzig 1882).

⁸⁷ Problem ten wpisuje się w szerszy spór na temat istoty dowodów matematycznych. Jak pokazuje Krzysztof Wójtowicz, spór ten nie jest izolowany od zagadnienia wyjaśniania w matematyce; zob. *ibidem*, *passim*.

⁸⁸ Cyt. za: R. Murawski, *Filozofia matematyki...*, op. cit., s. 211.

⁸⁹ Zob. J. Dadaczyński, *Filozofia matematyki w ujęciu historycznym*, op. cit., s. 95.

euklidesowa zyskała czy straciła na prostocie, jednak pewne jest, że zyskała większy stopień ścisłości. W tym miejscu kończę tę dygresję, choć zapewne zyskałaby ona, gdyby wzbogacić ją o szereg szczegółów technicznych. Moim zdaniem przykład syntaktycznego podejścia do geometrii jest głęboko pouczający. Pozostaje omówić teraz jeszcze jedną – psychologiczną – strategię wyjaśniania naukowego.

6.4. Psychologiczne strategie wyjaśniania

Zwolennicy omówionych wcześniej strategii pojmowali wyjaśnienie w kategoriach operacji na zdaniach (strategia syntaktyczna) lub odkrywania i reprezentowania obiektów i procesów (strategia semantyczna). Wiele osób pytanych: „co znaczy wyjaśnić określone zjawisko?” odpowiada jednak: „zrozumieć je” czy też „zrozumieć jego przyczyny”. Proces wyjaśniania można by więc utożsamiać z „ruchem myśli”, który zmierza do objęcia problemu intelektem, czyli zrozumienia. W takim wypadku wyjaśnianie staje się tożsame z *rozumieniem*.

Problem polega jednak na tym, że powszechnie używane pojęcie *rozumienia* jest niejasne. Trudno jest nawet wytłumaczyć, jak miałyby przebiegać wyjaśnianie oparte na rozumieniu. Bez wchodzenia w szczegóły można od razu stwierdzić, że strategia taka jest niezadowolająca i obciążona błędem *ignotum per ignotum*. Wyjaśnianie powinno być czymś intersubiektywnym i komunikowalnym w łatwy sposób, podczas gdy *rozumienie* jest ze swej natury subiektywne i niekomunikowalne. Ponadto omówione poprzednio koncepcje wyjaśniania naukowego miały wyraźne komponenty normatywne. Mówiły one, jakie warunki *powinno* spełniać wyjaśnianie naukowe. Poprzednie analizy wskazały, że warunkami tymi są Hempłowskie warunki (H1)–(H2) oraz sformułowane na podstawie paradoksów wyjaśniania warunki (W1)–(W5). Moim zdaniem rezygnacja z komponentu normatywnego byłaby destrukcyjna dla samej idei wyjaśniania naukowego. Podejście psychologiczne z całą pewnością nie jest nor-

matywne. Ze względu na ogromny stopień niejasności trudno jest nazwać je nawet deskryptywnym. Paul Churchland zapewne zaliczyłby pojęcie *rozumienia* do słownika „psychologii potocznej” (*folk psychology*), który powinien zostać wyeliminowany (na zasadzie redukcji eliminatywnej) i zastąpiony słownikiem neuronaukowym⁹⁰.

Psychologicznej strategii wyjaśniania można by bronić co najmniej na dwa sposoby. Po pierwsze można by się odwoływać do różnych koncepcji „filozofii humanistycznej”, takich jak Gadamerowskie „koło hermeneutyczne”. Problem polega jednak na tym, że koncepcja ta jest niejasna i wymaga interpretacji. Drugim sposobem obrony koncepcji psychologicznej jest odwołanie się do narzędzi stosowanych w szeroko rozumianej kognitywistyce. Wówczas mogłaby zostać skonstruowana neuropsychologiczna teoria rozumienia. Prawdopodobnie mogłaby ona oprzeć się niektórym z powyższych zarzutów, niekoniecznie jednak wszystkim. Teorię taką – utrzymaną w duchu koneksjonistycznym – zaproponował między innymi wspomniany wyżej Paul Churchland⁹¹. Jego zdaniem wyjaśnienie można utożsamić z odnalezieniem określonego wzorca aktywacji (*pattern of activation*) sieci konekcyjnej.

Zaproponowane przez Churchlanda podejście do wyjaśniania jest oparte na idei rozproszonego przetwarzania informacji w reżimie równoległym (*parallel distributed processing*, w skrócie: PDP). Ogólnie mówiąc, PDP to „przekształcanie jednego wzorca w inny podczas ich przechodzenia przez liczne skonfigurowane połączenia synaptyczne”⁹². Innymi słowy, na różnych etapach przetwarzania informacji przez mózg przesyłane wzorce podlegają różnym modyfikacjom. Zdaniem Churchlanda, PDP jest zasadą działania mózgu i magazynowania w nim informacji. PDP może być modelowane

⁹⁰ Zob. P.M. Churchland, *Eliminative Materialism and the Propositional Attitudes*, „Journal of Philosophy” 1981, no. 78, s. 67–90.

⁹¹ Zob. P.M. Churchland, *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*, The MIT Press, Cambridge, MA 1989; C.F. Craver, *Explaining the Brain...*, *op. cit.*, s. 28–34.

⁹² P.M. Churchland, *Mechanizm rozumu, siedlisko duszy...*, *op. cit.*, s. 20.

w sposób sztuczny – obecnie jest to jedna z wiodących strategii w ramach sztucznej inteligencji. Paul Churchland od wielu lat bada kodowanie różnych własności w abstrakcyjnych przestrzeniach za pomocą wektorów. W tym kontekście wyjaśnianie

(...) polega na aktywacji określonego prototypu wektora w dobrze wytrenowanej sieci. Polega to na ujęciu problematycznego przypadku jako przykładu ogólnego rodzaju, który posiada bogatą w informacje i szczegółową reprezentację⁹³.

Innymi słowy, proces wyjaśniania polega na dopasowaniu zjawiska do neuronalnego prototypu, czyli wzorca zakodowanego w mózgu⁹⁴. Kodowane są zarówno obiekty, jak i ich własności. Churchland traktuje wspomniane „wzorce aktywacji sieci” czy też „prototypowe wektory” w sposób realistyczny. Jego zdaniem różne typy rozumienia są powiązane z różnymi prototypami. Do prototypów wykorzystywanych w wyjaśnianiu Churchland zalicza między innymi: (1) prototypy zespołów własności, (2) prototypy następstw czasowych, (3) prototypy funkcjonalne, (4) prototypy nadrzędne (odnoszące się np. do praw geometrii), (5) prototypy interakcji społecznych, czy też (6) prototypy motywacji. Zapewne listę tę można by jeszcze rozszerzyć.

Psychologiczna strategia wyjaśniania oparta na PDP jest na pewno lepsza niż strategia wyjaśniania oparta na „psychologii potocznej”. Nie jest ona jednak wolna od problemów. Carl Craver zarzuca koncepcji Churchlanda, że nie spełnia ona norm nałożonych na wyjaśnianie naukowe⁹⁵. Przykładowo, warunek wyjaśniania (W1) głosi, że samo stwierdzenie następstwa czasowego zjawisk nie jest wystarczające, aby można było mówić o wyjaśnieniu. Churchland

⁹³ *Idem, A Neurocomputational Perspective...*, *op. cit.*, s. 210.

⁹⁴ Oczywiście termin *prototyp* jest tu używany w innym kontekście niż w przypadku prototypów, o których mowa w lingwistyce kognitywnej (wspominałem o nich w rozdziale 2). Churchland byłby jednak zapewne skłonny twierdzić, że każdy prototypowy egzemplarz kategorii (pojęcie) odpowiada prototypowemu wzorcowi aktywacji sieci neuronalnej.

⁹⁵ Zob. C.F. Craver, *Explaining the Brain...*, *op. cit.*, s. 31.

uznaje natomiast prototyp następstwa czasowego za wyjaśniający⁹⁶. Problem ten nie jest jednak – moim zdaniem – poważny, gdyż co do zasady można by sformułować warunki zabezpieczające przed nie-relevantnymi wyjaśnieniami. Według mnie istotniejsza jest jednak inna kwestia.

Jak już wspomniałem, wyjaśnienia naukowe powinny być intersubiektywne. Wymóg ten jest spełniony poprzez konstrukcje graficzne (schematy, diagramy, wykresy) reprezentujące PDP oraz odpowiednie wzorce aktywacji. Wówczas psychologiczna koncepcja wyjaśniania zbliża się do strategii semantycznej. W takiej interpretacji koncepcja Churchlanda nie wprowadza nowej jakości. Strategia psychologiczna nie może być jednak w łatwy sposób zredukowana do reprezentacji graficznych. Trudno jest zaprzeczyć Churchlandowi, że w wyjaśnianiu pewną rolę odgrywa komponent psychologiczny. Przykładowo można argumentować, że procesowi wyjaśniania towarzyszy zwykle budowanie spójnego modelu przekonań. Sądzą jednak, że przeciw psychologicznym strategiom wyjaśniania staje przytaczany już argument Karla Poppera. Filozofa nauki powinny interesować przede wszystkim wytwory procesów psychicznych, będące mieszkańcami Popperowskiego „trzeciego świata”.

6.5. W stronę wyjaśniania neuronaukowego

Przypomnijmy raz jeszcze argument Karla Poppera sformułowany w *Wiedzy obiektywnej*:

Prawdą jest, że działania lub procesy podpadające pod ogólny termin „rozumienie” są działaniami subiektywnymi, osobistymi lub psychologicznymi. Należy je odróżnić od (mniej lub bardziej udanych) *efektów* tej działalności, od ich rezultatów, od „ostatecznego stanu” rozumienia (w określonym czasie). (...) Z chwilą, gdy subiektywny stan

⁹⁶ Zob. P.M. Churchland, *A Neurocomputational Perspective...*, *op. cit.*, s. 213.

rozumienia zostanie w końcu osiągnięty, psychologiczny proces wiódący ku niemu powinien być analizowany w kategoriach przedmiotów trzeciego świata, z którymi jest związany. W istocie można go analizować *wyłącznie* w tych kategoriach⁹⁷.

Wytwory procesów psychicznych stają się elementami teorii naukowych, które można uznawać za zbiory zdań. Nie ma powodu, by traktować teorie neurokognitywne inaczej niż wszystkie inne teorie naukowe. Oczywiście nie oznacza to, że neuronauka poznawcza nie różni się niczym od np. fizyki czy kosmologii. Cechami specyficznymi teorii neurokognitywnych jest wysoki stopień uteoretyzowania badań, który jest związany z dużą „składową hermeneutyczną”, oraz kryteria wyboru najlepszych wyjaśnień, jakimi są koherencja i konwergencja. Mimo słabego zmatematyzowania neuronauki poznawczej kryteria te pozwalają w racjonalny sposób wybierać hipotezy i testować całe teorie.

Jak pokazałem w poprzednich podrozdziałach, zarówno semantyczne, jak i psychologiczne strategie wyjaśniania „kolapsują” do strategii syntaktycznej. Pozostaje jednak jeszcze zmierzyć się z problemami, w jakie są uwikłane strategie syntaktyczne (np. model D-N). Problemy te nie przekreślają w żadnym wypadku samej idei wyjaśniania jako operacji na zdaniach. Różne paradoksy pokazują tylko, że wyjaśnianie w stylu neopozytywistycznym jest „zbyt szerokie”. Eliminacja paradoksów jest możliwa przez sformułowanie pewnych warunków ograniczających ((W1)–(W5)). Warto zastanowić się, skąd biorą się te paradoksy. Wróćmy jeszcze na chwilę do najbardziej modelowej koncepcji wyjaśniania w stylu neopozytywistycznym. W koncepcji dedukcyjno-nomologicznej wyjaśnianie wygląda następująco:

$$\frac{L_1, L_2, \dots, L_r}{C_1, C_2, \dots, C_k} \\ E$$

⁹⁷ K.R. Popper, *Wiedza obiektywna*, *op. cit.*, s. 202–204.

Z przesłanek (*explanans*) w postaci przynajmniej jednego prawa, czyli zdania ogólnego, oraz zdań odnoszących się do warunków początkowych wyprowadzamy wniosek, którym jest *explanandum*. Schemat jest prosty. Pytanie brzmi: czy nie nazbyt prosty? Oczywiście, przesłanki (*explanans*) mogą składać się z bardzo dużej liczby zdań. Wyjaśnianie w stylu neopozytywistycznym jest najbardziej naturalną konsekwencją przyjęcia zdaniowej koncepcji teorii naukowej.

Teorie neuronaukowe są wysoce skomplikowanymi tworam. Mówiąc kolokwialnie, jeśli chodzi o „wielkość”, są one ponadto dość dużymi zbiorami zdań. Trudno wyobrazić sobie, aby w rzeczywistej praktyce naukowej, gdy porównywane są dziesiątki hipotez, a następnie konfrontowane są one z bardzo różnymi źródłami danych, wyjaśnienie miało polegać na wyliczeniu praw oraz warunków początkowych, a następnie przeprowadzeniu wnioskowania. Teoria nie jest bowiem luźnym zbiorem zdań, ale logiczną całością. Oznacza to, że tworzące ją zdania wchodzą między sobą w rozmaite związki inferencyjne. Wspomniana już w pierwszym rozdziale Popperowska analogia teorii naukowych i sieci wydaje mi się bardzo trafna⁹⁸. W przypadku neuronauki poznawczej zadaniem tych sieci nie jest tylko wyłapywanie empirycznych danych i prawidłowości. Muszą radzić sobie także z całą gamą presupozycji oraz nieweryfikowalnych założeń, przyjmowanych w ramach różnych paradygmatów neuronauki poznawczej.

Kilkukrotnie mowa była już o problemie praw w neuronauce poznawczej. Na gruncie wykorzystywanej przeze mnie siatki wnioskowania do najlepszego wyjaśnienia można bez wielkiej szkody zrezygnować z upodobnienia za wszelką cenę neuronauki poznawczej do nauk nomotetycznych, np. fizyki. Jednoznaczne jest to z ograniczeniem zastosowania koncepcji D-N jedynie do szczególnych przypadków, takich jak wyjaśnianie na gruncie psychologii ewolucyjnej poprzez odwołanie się do działania modułów. Równie dobrze można wskazać jednak przykłady, choćby takie jak depolaryzacja błony

⁹⁸ Zob. K.R. Popper, *Logika odkrycia naukowego*, op. cit., s. 53.

komórki nerwowej, gdy znacznie lepiej zastosować wyjaśnianie mechanistyczne. Prowadzi nas to do wniosku, że przyjęcie określonej strategii wyjaśniania naukowego jest zależne od interesującego nas w danym momencie poziomu teorii neurokognitywnej. José Luis Bermúdez wprowadził użyteczną dystynkcję na wyjaśnianie w sensie horyzontalnym i wertykalnym. Wspomniana wyżej relatywizacja do określonego poziomu teoretycznego oczywiście odnosi się do pierwszego przypadku. Co jednak z drugim, czyli wykazaniem łączności pomiędzy wyjaśnieniami partykularnych zjawisk na różnych poziomach? Aby rozwiązać ten problem, przyjrzymy się ponownie teorii ewolucji. Dlaczego spośród wielu teorii formułowanych w ramach różnych nauk przyrodniczych akurat darwinizm (wraz z jego późniejszymi uzupełnieniami) jest tak ważny?

Daniel Dennett przyznaje teorii ewolucji szczególny status. Pisze on: „Jeśli miałbym przyznać nagrodę za najlepszą ideę wszech czasów, dałbym ją Darwinowi przed Newtonem, Einsteinem i wszystkimi innymi”⁹⁹. Oczywiście wypowiedź ta ma w dużej mierze charakter emocjonalny. Dalej jednak Dennett w charakterystyczny dla siebie, barwny sposób uzasadnia swoje przekonanie: „Koncepcja ewolucji przez dobór naturalny za jednym zamachem jednoczy sferę życia, sensu i celu ze sferą przestrzeni i czasu, przyczyny i skutku, mechanizmów i praw fizyki”¹⁰⁰. W podobnym duchu są utrzymane słynne słowa Theodosiusa Dobzhansky’ego: „Nic w biologii nie ma sensu, jeśli jest rozpatrywane w oderwaniu od ewolucji”¹⁰¹. Teoria ewolucji *unifikuje* bardzo dużą liczbę danych z zakresu nauk biologicznych. Tworzy ona jednolity *wzorzec*, pod który podpadają np. cytologia, dendrologia, bakteriologia, ornitologia itd. Dzięki teorii ewolucji wyniki uzyskane w ramach tych – oraz rzeszy innych – dziedzin przyczyniają się do wyjaśnienia licznych zagadnień przyrodniczych. Istotna

⁹⁹ D.C. Dennett, *Darwin’s Dangerous Idea. Evolution and the Meanings of Life*, Simon & Schuster Paperbacks, New York–London 1995, s. 21.

¹⁰⁰ *Ibidem*.

¹⁰¹ T. Dobzhansky, *Nothing in Biology Makes Sense Except in the Light of Evolution*, „The American Biology Teacher” 1973, vol. 35, no. 3, s. 125–129.

koncepcja wyjaśniania została opracowana przez Philipa Kitchera¹⁰². Jego zdaniem wyjaśnianie polega na *unifikacji*. Dodam, że prócz teorii ewolucji za paradygmaticzny przykład teorii unifikującej uznaje on mechanikę klasyczną. Koncepcja unifikacyjna należy do zdaniowych strategii wyjaśniania naukowego. Według Kitchera przed paradoksami wyjaśniania w stylu neopozytywistycznym bronią nas wzory rozumowania (uzasadniania) tworzone przez teorie unifikujące¹⁰³. Jak pisze Kitcher:

(...) teoria unifikuje nasze przekonania, gdy dostarcza jednego (lub ogólniej: niewielu) wzoru(ów) rozumowania, który może być użyty w wyprowadzeniu wielu zdań, które akceptujemy¹⁰⁴.

Kitcher twierdzi, że teorie unifikujące tworzą wzory rozumowań. Podkreśla on, że akceptacja pewnej przesłanki, podpadającej pod określony wzór rozumowania, zobowiązuje automatycznie do akceptacji innych przesłanek¹⁰⁵. Wzory rozumowań powinny być stosowalne do wyjaśniania jak największej liczby zjawisk. Jego zdaniem:

Nauka posuwa naprzód nasze rozumienie natury, pokazując nam, jak derywować opisy wielu zjawisk, wykorzystując ten sam wzór derywacji jeszcze raz i jeszcze raz. Pokazując to, nauka uczy nas, jak redukować liczbę faktów, które musimy zaakceptować jako ostateczne¹⁰⁶.

Jak widać, w powyższych cytatach pojawiają się terminy psychologiczne („przekonania”, „rozumienie”). Zdaniem Kitchera,

¹⁰² Zob. Ph. Kitcher, *Explanatory Unification*, „Philosophy of Science” 1981, no. 48, s. 507–531.

¹⁰³ Zob. *ibidem*, s. 522–526.

¹⁰⁴ *Ibidem*, s. 514.

¹⁰⁵ Zob. *ibidem*, s. 515–519.

¹⁰⁶ Ph. Kitcher, *Explanatory Unification and the Causal Structure of the World*, [w:] *Scientific Explanation*, red. Ph. Kitcher, W.C. Salmon, The University of Minnesota Press, Minneapolis 1989, s. 410–505.

wyjaśnianie ma komponent psychologiczny, jednak na nim się nie kończy. W ostatecznym rozrachunku rozumowania wyjaśniające przeprowadzane są na zdaniach. Co istotniejsze na tym etapie, koncepcja Kitchera wiedzie nas do ogólniejszego wniosku, zgodnie z którym wyjaśnienia naukowe opierają się na całej sieci związków wewnątrz teorii. Zatem „izolowane” rozumowania w stylu neopozytywistycznym nie są dla niego satysfakcjonującymi wyjaśnieniami. Nie oznacza to jednak, że należy odrzucić całkowicie koncepcję wyjaśniania, w której przesłankami są prawa. Każde rozumowanie wyjaśniające jest wbudowane w większą strukturę argumentacyjną. Innymi słowy, każde rozumowanie unifikacyjne opiera się na pewnej *prawidłowości*. Tyle tylko, że prawidłowości te nie muszą być uznawane za „zewnętrzne” wobec teorii, np. za „odwieczne” *prawa przyrody*. Prawidłowości, o których mowa, są tworzone raczej przez sieć związków pomiędzy poszczególnymi zdaniami teorii.

Uważam, że koncepcję Kitchera można z powodzeniem zastosować w przypadku teorii neurokognitywnych. Podstawowy problem polega jednak na wskazaniu konkretnych teorii, które mogą pełnić rolę teorii unifikujących. W związku ze specyfiką neuronauki poznawczej nie jest możliwe, przynajmniej na razie, wskazanie tak eleganckich i bogatych w treść empiryczną teorii unifikujących, jak np. Newtonowska mechanika klasyczna. Sądzę jednak, że w przypadku neuronauki poznawczej analogiczną rolę pełnią paradygmaty, takie jak paradygmat komputerowy, psychologia ewolucyjna oraz *embodied-embedded mind*. Jeśli mam rację, problem wyjaśniania neuronaukowego jest ściśle połączony z zagadnieniem jedności nauki. Jak już pisałem w poprzednim rozdziale, ideał jedności nauki może być ocalony na gruncie ujęcia sieciowego. W tym kontekście przeanalizowałem krótko koncepcję teorii międzyzakresowych stworzoną przez Darden i Maull, ilustrując ją przykładem *Neural Engineering Framework* (NEF) Eliasmitha¹⁰⁷. Zastosowanie teorii międzyzakresowych

¹⁰⁷ Zob. M. Miłkowski, *Theoretical Unification and the Neural Engineering Framework*, *op. cit.*

pokazuje ponadto, jakiego rodzaju rozumowania wchodzą w grę, gdy analizujemy wyjaśnianie naukowe w sensie wertykalnym.

Rekapitułując: w przypadku neuronauki poznawczej wybór najlepszej strategii eksplanacyjnej jest uzależniony od tego, na jakim poziomie teoretycznym znajduje się interesujące nas zagadnienie. W przypadku niższych poziomów, takich jak pojedyncze neurony oraz złożone struktury, najbardziej optymalne wydaje się wyjaśnianie mechanistyczne, które korzysta nie tylko z opisu werbalnego, lecz także z reprezentacji graficznych. Z kolei w przypadku najwyższego poziomu, czyli procesów poznawczych, z powodzeniem można stosować syntaktyczną koncepcję wyjaśniania, której najbardziej modelowym przypadkiem jest koncepcja dedukcyjno-nomologiczna. Argumentowałem jednak, że wybór strategii wyjaśniania jest kwestią ekonomii poznawczej. Wszystkie typy wyjaśniania sprowadzają się bowiem na fundamentalnym poziomie do operacji na zdaniach. Operacje te nie są jednak izolowane, lecz zawsze wpisują się w szerszą strukturę argumentacyjną.