

Marcin Miłkowski

Filozofia jako inżynieria odwrotna: rzecz o naturalizmie Daniela C. Dennetta

We współczesnej filozofii analitycznej coraz wyraźniej zaznacza się naturalizm, pojmowany jako program uprawiania i zmiany samej filozofii. Postawę naturalistyczną przyjmuje wielu filozofów, zmierzających do naturalizacji umysłu i to naturalizacji na modłę wyraźnie inżynierską. Jednym z nich jest Daniel C. Dennett:

Moją fundamentalną perspektywą jest *naturalizm*, czyli idea, że rozważania filozoficzne nie są wyższe ani pierwotniejsze od rozważań w ramach nauk przyrodniczych, lecz przebiegają we współpracy z tymi sposobami poszukiwania prawdy, i że właściwym zadaniem filozofów jest wyjaśnianie i integracja często walczących ze sobą perspektyw w jedną wizję wszechświata (Dennett [2003a]), s. 14–15).

Współpracę tę Dennett pojmuje dosłownie: filozofowie są partnerami naukowców prowadzących badania empiryczne, czyli doradzają im, jak na przykład winna wyglądać architektura umysłu homonoidalnego robota (por. Dennett [1998], s. 153–170). Do roli doradcy filozof się nie ogranicza. Stawia też problemy rozwiązywalne przez nauki empiryczne¹, ale mające konsekwencje dla tradycyjnej problematyki filozoficznej. Andrzej Klawiter nazywa problemy stawiane przez filozofów umysłu w kognitywistyce „problemami kognitywnymi”; problem kognitywny „jest zgodny z zasadniczym korpusem wiedzy w danym obszarze oraz możliwe jest definitywne rozstrzygnięcie tego problemu przy wykorzystaniu narzędzi badawczych powszechnie uznawanych za efektywne w tej dziedzinie” (Klawiter [1999], s. 137). Problemy kognitywne nie są więc już problemami, które mogłaby rozwiązać analiza języka potocznego czy idealnego języka nauki, fenomenologia czy ontologia formalna. Tego rodzaju problemy pozostają przecież poza polem ich zainteresowań.

¹ W wypadku Dennetta przez „nauki przyrodnicze” rozumieć należy raczej szeroko wszelkie nauki empiryczne.

Uzyskanie spójnej, racjonalnej wizji świata – a więc także umysłu – wymaga wyjaśnienia, w jaki sposób umysł funkcjonuje. Umysł rozumiany nie tylko i nie przede wszystkim biologicznie czy neurofizjologicznie, lecz w kategoriach intencjonalnych. W perspektywie naturalistycznej wyjaśnienie to dotyczy więc związku między intencjonalnym umysłem a dostępnymi empirycznie w naukach przyrodniczych zjawiskami. Naturalista nie może ograniczyć się do zadeklarowania, że związek taki zachodzi, gołosłownie twierdząc, że problem psychofizyczny został rozwiązany. Musi raczej pokazać *wyjaśnienie* takiego związku. Uzasadnione wydaje się więc nazwanie tego stanowiska *naturalizmem eksplanacyjnym*:

(NE) Wszystkie zjawiska empiryczne można wyjaśniać zgodnie z naukami przyrodniczymi: albo (1) w ramach owych nauk, albo (2) stosując metody zgodne ze standardami racjonalności naukowej.

Teza (NE) nie głosi, że nauki przyrodnicze wszystko efektywnie wyjaśniają², lecz że wszystko, co się da wyjaśnić, da się wyjaśnić albo w ramach owych nauk, albo zgodnie ze standardami racjonalności naukowej, np. przez racjonalną argumentację czy eksperyment myślowy. Naturalistyczne wyjaśnienie danego zjawiska polega na znalezieniu mechanizmu, który odpowiada za istnienie tego zjawiska, przy czym musi to być mechanizm prostszy od wyjaśnianego zjawiska i prawdopodobny w świetle współczesnej nauki. Szczegóły wyjaśnień naukowych są często nieistotne dla filozofów, wystarcza fakt, że takie wyjaśnienia istnieją³.

W jaki sposób można nadać prawomocność tezie (NE)? Najprościej – rozumieć ją jako pewien program i po prostu zgodnie z naukami przyrodniczymi (np. korzystając z dotychczasowych wyników nauki) wyjaśniać kolejne zjawiska, w tym zwłaszcza stany intencjonalne i przeżycia świadome. Naturalizm eksplanacyjny staje się tezą tym bardziej uzasadnioną, im więcej potwierdzono empirycznych hipotez dotyczących funkcjonowania umysłu, przeto filozofia wymaga tu ścisłej współpracy z nauką. Jeśli sama nauka jeszcze takich mechanizmów nie zna, naturalistyczny filozof często jest gotów sformułować hipotezy empiryczne na ich temat. Dennetta wyróżnia przywiązanie do nauk inżynierskich, przez które rozumie nie tylko przedmioty wykładane na politechnikach, lecz także biologię ewolucyjną i kognitywistykę pojętą jako tzw. inżynierię odwrotną (*reverse engineering*). Inżynieria odwrotna to proces analizy gotowego urządzenia przez inżyniera, który chce poznać sposób jego działania i funkcje poszczególnych elementów.

Niżej pokażę najpierw, jak działa inżynieria odwrotna w biologii i kognitywistyce. Następnie przedstawię przykład inżynierskiego rozumowania, podważają-

² Stąd aspekt niedokonany czasownika „wyjaśniać” w tezie (NE).

³ Wyjątkiem jest sytuacja, gdy jakieś szczegóły wyjaśnienia podważają tradycyjną tezę filozoficzną; np. wyniki astronomii podważają aprioryczną dedukcję liczby planet w Układzie Słonecznym.

cego tradycyjne rozumienie cech świadomości w filozofii. Okaże się też, że za wyjaśnienia inżynierijne Dennett jest gotów płacić dosyć wysoką cenę.

1. Inżynieria Matki Natury

Dennett zalicza biologię do nauk inżynierskich (Dennett [1995], s. 187). Inżynierska perspektywa jest czymś, bez czego nie można uporządkować myśli darwinowskiej i co stanowi źródło jej siły. Biolog ma inny cel od zwykłego inżyniera: nie idzie mu o stworzenie nowego urządzenia, lecz tylko o odtworzenie jego konstrukcji, zrozumienie jego budowy. Dokonuje więc inżynierii odwrotnej, znanej też zwykłym inżynierom (Dennett [1995], s. 212).

Na czym polega inżynieria odwrotna? Najłatwiej opisać to na przykładzie. Elektronik chce dowiedzieć się, jaką budowę ma pewien świetny wzmacniacz skonstruowany przez jego konkurenta. Rozkręca go więc i zakładając, że konkurent wiedział, co robi – to jest stosował rozwiązania racjonalne, optymalne ze względu na jakość dźwięku wzmacniacza – próbuje rozpoznać odpowiednie funkcjonalne składniki wzmacniacza: zasilacz, wzmacniacz właściwy (np. tranzystorowy), potencjometr i tak dalej. Następnie bada, jak te elementy zostały skonstruowane; sprawdza, jakiego rodzaju przewodów i połączeń użyto, jakiej marki podzespołów... Jak powiada Dennett, bez założenia optymalności nie można daleko ujechać, choć bywa ono fałszywe, gdyż nikt nie tworzy od razu urządzeń idealnych. Jednak rozpoznanie bloków funkcjonalnych urządzenia bez założenia racjonalności konstruktora, czyli optymalności budowy, nie jest w ogóle możliwe.

Nieco inaczej wygląda inżynieria odwrotna wytworów ewolucji. W świecie przyrody racjonalnym konstruktorem tych wytworów jest wedle Dennetta dobór naturalny, personifikowany jako obdarzona intencjonalnością Matka Natura⁴. Próbujemy mianowicie zgadnąć, jaką ewentualną rację „odkryła” lub „wyróżniła” Matka Natura, tworząc takie a nie inne obiekty (Dennett [1995], s. 213). Matka Natura postępuje jednak nieco inaczej od inżynierów-ludzi:

Matka Natura, gdy tworzy jakiś organizm, to zawsze na zasadzie: „Kupą, mości panowie! Im nas więcej, tym weselej! Zabawimy się i jakoś to będzie” (Dennett [1996], s. 253).

Funkcje organów zmieniają się w czasie, ale organ zmieniający funkcję niekoniecznie zanika – nie jest wymieniany na nowy, sprawniejszy element, lecz zaczyna realizować kolejną funkcję, następnie jeszcze jedną i tak dalej. Powstają w ten sposób konstrukcje bardzo złożone i trudne do analizy, a o ich optymalno-

⁴ Personifikacja ta budzi liczne wątpliwości u plemistów Dennetta; tutaj na mocy zasady życzliwości w interpretacji zakładam, że wątpliwości te da się w jakiś sposób rozwiązać. Założenie to traktuję jako robocze.

ści w danym środowisku świadczy to, że tak skonstruowane organizmy (jeszcze) nie wymarły.

Podstawą inżynierii odwrotnej u Dennetta jest teoria systemów intencjonalnych. W teorii tej obiektem intencjonalnym jest każdy taki obiekt, którego aktywność opłaca się przewidywać z użyciem słownictwa intencjonalnego (Dennett [2003b])⁵. Ową opłacalność mierzy się złożonością opisu: obiekty można opisywać za pomocą terminologii czysto fizycznej, lecz opisy stają się niezwykle skomplikowane, gdy mają służyć do poprawnego przewidywania racjonalnych działań ludzkich. W czysto fizycznych kategoriach trudno też przewidywać działanie układów funkcjonalnych takich jak zegary czy komputery. Te najlepiej opisuje się i przewiduje za pomocą terminologii funkcjonalnej. Wykorzystywanie słownictwa intencjonalnego do przewidywania i wyjaśniania zachowania obiektów Dennett określa mianem przyjmowania nastawienia intencjonalnego (przyjmować można też nastawienie funkcjonalne oraz fizyczne).

Taka koncepcja intencjonalności umożliwia przeformułowanie rozumowań dotyczących adaptacji organizmów do środowiska. Weźmy pewnego żerującego w nocy ssaka, wyposażonego w uszy czułe na ultradźwięki i narząd zdolny do wydawania ultradźwięków. Szukając adaptacyjnej funkcji tych narządów, stwierdzamy, że najprawdopodobniej służą do pobierania informacji ze środowiska na zasadzie echolokacji. Ten wniosek jest *zawodny* – „Matka Natura” nie jest idealnym systemem intencjonalnym i nie wszystko robi w sposób całkowicie racjonalny – lecz nikt nie neguje, że nietoperze posługują się echolokacją.

Zdaniem Dennetta odwrotna inżynieria – poszukiwanie racjonalnych uzasadnień w postaci wartości adaptacyjnej dla replikatora – jest podstawą całej biologii ewolucyjnej (Dennett [1995], s. 238). Podobne stanowisko zajmuje Steven Pinker [2002] i jeden z najwybitniejszych biologów ewolucyjnych John Maynard Smith [1995]. Idea odwrotnej inżynierii w biologii dociera do najbardziej poczytnych czasopism naukowych (np. poświęcony jej artykuł Csete i Doyle’a [2002] opublikowano w piśmie „Science”). Jednak doczekała się do tej pory niewielu analiz filozoficznych (wyjątkami są Lewens [2002] i Richardson [2003]).

Adaptacjonistyczna odwrotna inżynieria opisana powyżej prowadzi do wykrycia optymalnej funkcjonalnej struktury. Pojęcie funkcji wykorzystywane w tym celu może być rozumiane dwojako: (1) jako zwykła dyspozycja do działania (Cummins [1975]); (2) dyspozycja do działania istniejąca dlatego, że prowadzi ona do takiego a nie innego działania (Wright [1973]). Na podstawie stosowanego w ramach inżynierii odwrotnej pojęcia funkcji można wyróżniać różne rodzaje tej inżynierii.

⁵ Takie sformułowanie może budzić podejrzenia, iż stanowisko Dennetta jest instrumentalistyczne, choć on sam uważa, że jest umiarkowanym realistą co do zjawisk intencjonalnych. Sprawa ta wykracza jednak poza ramy tego artykułu.

Wyżej podawałem przykłady inżynierii oparte na analizie funkcjonalnej w sensie Cumminsa. Cummins pojęcie funkcji definiował jako zdolność wyróżnianą przez rolę przyczynową w ramach analizy zdolności (dyspozycji) całego układu (Cummins [1975], s. 765). W takiej zaś odwrotnej inżynierii chodzi o zidentyfikowanie, które części i w jaki sposób składają się na dyspozycje do działania całego układu, czyli analizowanego organizmu. Jest to aposterioryczne odtwarzanie funkcjonalnej konstrukcji organizmu na podstawie jego obserwowanej struktury fizycznej. Odwrotna inżynieria może brać pod uwagę środowisko i korzystać z pewnych apriorycznych założeń⁶. Otrzymujemy przeto podział (Richardson [2003], s. 1282):

	analiza organizmu	analiza środowiska
<i>a priori</i>	wnioskowanie na podstawie ogólnych fizycznych zasad konstrukcji	wnioskowanie na podstawie zasad fizycznych określających wymogi środowiska
<i>a posteriori</i>	wnioskowanie o konstrukcji funkcjonalnej organizmu na podstawie jego obserwowanej struktury	wnioskowanie o wymogach środowiska na podstawie obserwowanych efektów doboru naturalnego

Wymogi środowiska w biologicznej odwrotnej inżynierii traktuje się jako problemy, które w ewolucji ma rozwiązać dobór naturalny. Niezwykle trudno jednak zidentyfikować problemy niezależnie od ich rozwiązań i bez historycznych warunków początkowych (Griffiths [1996], Godfrey-Smith [1998a], s. 713–714). Przeciwnicy inżynierskiego adaptacjonizmu odrzucają go często dlatego, że negują możliwość wyróżnienia dobrych rozwiązań inżynierskich w oderwaniu od historycznych warunków rozwoju organizmów. Dennett taką możliwość uznaje – zbiór wszystkich możliwych biologicznie funkcjonalnych konstrukcji organizmów nazywa „przestrzenią konstrukcyjną” (*design space*): abstrakcyjną przestrzenią wielowymiarową, na którą odwzorowywane są różne rodzaje funkcjonalnych konstrukcji organizmów.

Na głębsze powiązanie z historycznymi uwarunkowaniami pozwala użycie bardziej etiologicznego pojęcia funkcji w sensie Larry’ego Wrighta [1973] (por. Allen i Bekoff [1995], Amundson i Lauder [1998], s. 230–231 i Godfrey-Smith [1998b]) czy też funkcji właściwej w sensie Millikan [1984]. Funkcja w sensie Wrighta jest definiowana w następujący sposób:

To, że funkcją *X*-a jest *Z*, oznacza, iż:

(a) *X* istnieje, ponieważ wykonuje *Z*;

(b) *Z* jest następstwem (lub wynikiem) istnienia *X*.

⁶ Mogą być to założenia o charakterze heurystycznym czy regulatywnym; wiadomo na przykład, że urządzenia bardziej modułowe są odporniejsze na uszkodzenia.

Jest to bardziej etiologiczne i teleologiczne pojęcie funkcji niż pojęcie Cumminsa. Odnosi się ono tylko do *aktualnej* funkcji Z, a definicja pojęcia funkcji abstrahuje od całej historii ewolucyjnej danej funkcji. Millikan posługuje się podobnym pojęciem funkcji, co Wright, lecz bardziej podkreśla historię ewolucyjną powstających funkcji (Millikan [1984], s. 93). W ramach odwrotnej inżynierii można posługiwać się zarówno aktualistycznym pojęciem funkcji Wrighta, jak i bardziej historycznym ujęciem *funkcji właściwej* Millikan. Wówczas faworyzowane w ramach historii ewolucyjnej funkcje danego organu okazują się nieprzypadkowymi funkcjami – nie każda dyspozycja układu okazuje się jego funkcją, a tylko taka dyspozycja, która pojawia się zawsze i tylko wtedy, gdy istnieje odpowiedni funkcjonalny organ/moduł/element X^7 .

Nastawienie intencjonalne usprawiedliwia praktykę inżynierii odwrotnej. Inżynieria odwrotna jest skuteczna, gdyż dobór naturalny potrafi generować złożone konstrukcje funkcjonalne; niektóre konstrukcje funkcjonalne z kolei kwalifikują się same jako systemy intencjonalne. W ten sposób w wizji Dennetta „Matka Natura” staje się nie tylko hipotetycznym inżynierem złożoności ewolucyjnej, ale i źródłem intencjonalności.

2. Inżynieria robotów umysłowych

Inżynierię odwrotną Dennett odnosi także do ludzkiego umysłu:

Materializm współczesny – a przynajmniej moja jego wersja – akceptuje z radością twierdzenie, że *my* jesteśmy pewnego rodzaju robotami – złożonymi z robotów złożonych z robotów. Myślenie w kategoriach robotów to użyteczne ćwiczenie, gdyż usuwa ono wymówkę, że nie wiemy jeszcze wystarczająco wiele o mózgach, aby powiedzieć, że to, co *właśnie* się dzieje, może być istotne. Wymówka taka dopuszcza pewien rodzaj mętnego romantyzmu w odniesieniu do tajemniczych mocy mózgów [...]. Jeśli materializm jest prawdziwy, powinno być („w zasadzie”) możliwe zbudowanie materialnej rzeczy – nazwijmy ją mózgiem robota, która robi to samo, co mózg, a zatem egzemplifikuje tę samą teorię przeżywania, co nasza (Dennett [2003c]).

Związek między analizą funkcjonalną a odwrotną inżynierią jest ścisły, nie dziwi więc, że z inżynierii takiej w filozofii umysłu korzystają zwolennicy funkcjonalizmu. Teza funkcjonalizmu głosi bowiem, że charakterystyka umysłu powinna być podawana w kategoriach funkcjonalnych, a nie czysto fizycznych. Przez funkcjonalne elementy rozumieją tacy funkcjonałiści najczęściej pewne moduły obliczeniowe, gdyż przyjmują też przekonanie, iż umysł działa przede wszystkim jako urządzenie przetwarzające informacje, co sprowadza się praktycznie do tezy

⁷ Ścisłą definicję podają Allen i Bekoff [1995], s. 615, lecz jest ona zawężona tylko do funkcji biologicznych.

o jego obliczeniowym charakterze⁸. Zanalizowane jednostki funkcjonalne – moduły – umysłu stanowią jego konstrukcję, jego architekturę funkcjonalną.

Nawet jeśli odwrotna inżynieria przyniesie w rezultacie pewną analizę funkcjonalną, która odpowiada działaniu istniejącego urządzenia, niekoniecznie jest to analiza odpowiadająca intencjom ewentualnego twórcy (zespołu twórców; doboru naturalnego). Analiza ta wskazuje pewne *wystarczające*, ale nie *konieczne* warunki pojawienia się takich lub innych modułów funkcjonalnych. Jest to więc rodzaj wnioskowania zawodnego, acz przydatnego. Odwrotną inżynierię uprawia też wielu kognitywistów, jak np. Steven Pinker [2002].

Tego rodzaju rekonstrukcję inżynieryjną można by krytykować na prostej podstawie: wiedzieć, jak coś zrobić, a znać teorię – to zupełnie inne rzeczy. Mechanik samochodowy nie musi znać zasad działania katalizatora do silnika spalinyowego, aby go skutecznie naprawiać. W filozofii naturalistycznej jednak nie ma mowy o naprawianiu silników, lecz chodzi o tworzenie pewnej koncepcji umysłu; *koncepcji*, a nie *teorii* – uroszczenia filozofów naturalistycznych zwykle nie sięgają aż tak daleko. Koncepcja ta powstaje przez stawianie hipotez o strukturze funkcjonalnej umysłu; hipotezy te mogą być testowane w ramach kognitywistyki i może się okazać, iż proponowana architektura umysłu jest po prostu technicznie wadliwa. Tak więc inżynieryjny aspekt nie oznacza tutaj wulgarnego pragmatyzmu; nie idzie o jak najszybsze stworzenie artefaktu, któremu można by przypisać choć krztynę zdolności umysłowych, mimo że jego zasada działania byłaby niezrozumiała. Podobnie biologia ewolucyjna, korzystająca z metod analizy inżynieryjnej, nie staje się przez to nauką mniej teoretyczną, lecz po prostu związaną z pewnymi heurystycznymi sposobami poszukiwania wyjaśnień.

3. Model wielokrotnych szkiców

Model świadomości, jaki proponuje Dennett, jest wynikiem inżynierii odwrotnej. Jest empiryczny i ma tłumaczyć obserwowalne zjawiska świadome. Na jego przykładzie wyraźnie widać wady i zalety inżynieryjnego podejścia do umysłu. Inżynieria odwrotna nie jest w stanie bowiem nigdy zagwarantować, że uzyskana na jej drodze hipoteza zawiera konieczne warunki pojawienia się struktury, która podlega analizie inżynierskiej; hipoteza taka może zawierać co najwyżej warunki wystarczające.

Standardowe ujęcie świadomości zakłada, że istnieje jej strumień, składający się z kolejnych świadomych reprezentacji. Kolejne zdarzenia postrzegane przez podmiot są przedstawiane w świadomości w porządku odpowiadającym faktycz-

⁸ Każdy układ, w którym można wyróżnić wyjście generujące przynajmniej jedną wartość dyskretną, generowane w odpowiedzi na wejściowe dane dyskretne (lub zbiór pusty), można określić jako układ realizujący obliczenie cyfrowe, to znaczy komputer cyfrowy. Por. Cutland [1980].

nemu zachodzeniu tych zdarzeń. Kiedy na przykład patrzę na jeża, który wchodzi z polanki do lasu, najpierw w mojej świadomości pojawia się jeż stojący jeszcze na polance, a dopiero potem jeż w lesie. Tego rodzaju uporządkowanie wydaje się dla zdroworozsądkowej intuicji czymś oczywistym. Istnieją jednak eksperymenty, które przy tak rozumianym strumieniu świadomości wydają się co najmniej anomaliami.

Zjawiska ϕ (zwane też efektami stroboskopowymi) wiążą się z postrzeganiem pozornego ruchu. Umożliwiają one m.in. oglądanie filmów — postrzegamy płynny ruch, choć w istocie wyświetlane są po kolei nieruchome klatki. Nelson Goodman referuje kilka eksperymentów dotyczących tych zjawisk w książce *Jak tworzymy świat* (Goodman [1997], s. 88–102), a także przedstawia swoją interpretację:

Najprostsze i najlepiej znane zjawisko ruchu pozornego zachodzi przy dwukrotnej, dokonanej z 10–45-milisekundowym interwałem czasowym i z niewielkim interwałem przestrzennym [1,4° lub 4° – przyp. M.M.] ekspozycji plamki na kontrastowym tle. Gdy odstęp czasowy jest mniejszy, postrzegamy jednocześnie wystąpienie dwu plamek; gdy przerwa jest dłuższa, widzimy dwa kolejne błyski; jeśli jednak odstęp mieści się we wspomnianych granicach, widzimy jedną plamkę przemieszczającą się z jednej pozycji do drugiej (Goodman [1997], s. 88).

Goodman zaproponował psychologowi badającemu to zjawisko, Paulowi Kolarsowi, aby plamki były różnych barw. Rezultaty okazały się zaskakujące:

Kiedy, oczywiście z zachowaniem naszych interwałów czasoprzestrzennych, pokazać komuś najpierw czerwony kwadrat, a potem duże zielone (albo różowe) koło, to zobaczy on, że kwadrat, podczas gdy gładko przesuwają się, rozrasta i przekształca w koło, *aż do punktu mniej więcej w pół drogi, jest cały czas czerwony, po czym nagle staje się zielony (bądź różowy)* (Goodman [1997], s. 102).

Goodman pyta, jak to możliwe, że w wypadku „ruchu pozornego uzupełniamy pośrednie czasoprzestrzenne położenia plamki wzdłuż pewnej linii zaczynającej się tam, gdzie wystąpił pierwszy błysk, a kończącej się tam, gdzie wystąpi drugi, *zanim jeszcze ten drugi ma miejsce?* Skąd wiemy, w jaką to ma iść stronę?” (Goodman [1997], s. 90). Jest zwolennikiem hipotezy „retrospektywnej”: wrażenie ruchu tworzymy dopiero po spostrzeżeniu drugiego błysku i rzutujemy w przeszłość. Zdaniem Kolarsa zaś „uzupełnianie zachodzi w czasie rzeczywistym” i „kierunek uzupełniania jest antycypowany na gruncie poprzednich doświadczeń” (Goodman [1997], s. 99). Sęk w tym, że zjawisko powtarza się też wówczas, gdy podmioty badane nie miały żadnych innych tego rodzaju doświadczeń!

Dennett twierdzi, że po pierwsze nie zachodzi żadne uzupełnianie, ani retrospektywne, ani antycypujące; a po drugie – hipotezy korekty retrospektywnej i korekty antycypującej są pod względem eksperymentalnym nierozróżnialne (Den-

nett i Kinsbourne [1992]). O uzupełnianiu można mówić tylko wtedy, gdy akceptuje się istnienie teatru kartezyjańskiego, na którego scenie pojawia się strumień świadomości, który jest uzupełniany (Dennett [1992b]). Z inżynierskiego punktu widzenia obserwowane zjawisko wymaga postulowania innej struktury, gdyż uzupełnianie wydaje się bardzo mało wiarygodną hipotezą: trudno wskazać ewolucyjny sens uzupełniania, to znaczy pozytywną wartość adaptacyjną dla organizmów, w których takie uzupełnianie zachodzi. Aparatura umożliwiająca zajście uzupełnienia – czy też tworzenia hipotez antycypujących – jest bardzo rozbudowana i byłaby nieskuteczna.

Antycypowanie wymagałoby tworzenia hipotez na temat dalszego przebiegu zdarzeń; jednak z zaobserwowanym stanem rzeczy zgodnych jest bardzo wiele kolejnych hipotez, a przeszłe doświadczenie, jak wskazywał już Goodman, nie może stanowić ich źródła. Czemu mózg miałby z góry projektować przebieg obserwowanych zdarzeń i na jakiej właściwie podstawie udaje mu się w normalnych warunkach – kiedy tajemnicze efekty stroboskopowe nie występują – poprawnie wybierać hipotezy? Trudno wyobrazić sobie mechanizm, który byłby w stanie generować niemal wyłącznie poprawne hipotezy na podstawie samej tylko informacji o poprzednio zaobserwowanych zdarzeniach.

Z kolei retrospektywna korekta wymagałaby cenzurowania powstających reprezentacji, zanim staną się one świadome. Do świadomości docierałyby one z pewnym opóźnieniem, a w tym czasie mechanizmy nieświadome korygowałyby utworzone już – prawdopodobnie na podstawie antycypacji – obrazy. One byłyby też następnie „wyświetlane” na ekranie świadomości.

Zbudowanie takiej skomplikowanej maszynierii obliczeniowej wydaje się jednak możliwe. Wszak robot wyposażony w czarno-biały skaner do kodów kreskowych może mieć program, który rozpoznaje tylko kody kreskowe; każdą zatem linię usiłuje kategoryzować jako element kodu (w ten sposób stawia hipotezy antycypujące). Równie dobrze w razie napotkania niezgodności hipotezy z kolejnymi danymi napływającymi ze skanera mógłby korygować uprzednio wytworzoną kategoryzację i dopiero wówczas wyświetlałby – powiedzmy – liczbę zakodowaną na napotkanym kodzie kreskowym. Zdaniem Dennetta jednak imputowanie świadomości tego rodzaju mechanizmu pociąga za sobą konieczność przypisania jej ogromnych mocy obliczeniowych, gdyż przetwarzanie informacji wzrokowych bardziej skomplikowanych niż proste kody kreskowe wymagałoby znacznie większych inwestycji w obliczenia. To jest jednak hipoteza inżynierska: Matka Natura zwykle wybiera rozwiązania optymalne; mechanizm antycypacji czy korekt retrospektywnych byłby nieoptymalny, a więc świadomość nie zawiera tego mechanizmu. Jednak Dennett podkreśla, że Matka Natura *nie zawsze* działa optymalnie.

Charakter reprezentacji mentalnych i ich porządek czasowy stanowi podstawę argumentacji we wspólnym artykule Dennetta i psychologa Marcela Kinsbourne’a [1992]. Piszą, że nieprawdziwe jest przekonanie, iż kolejność nadcho-

dzenia bodźców percepcyjnych odpowiada kolejności występowania zjawisk w świadomości. Reprezentacje bodźców nie są bowiem przetwarzane szeregowo, lecz współbieżnie – sieci neuronowe już u tak prostych ssaków naczelnych jak makaki mają 36 ośrodków wzrokowych przetwarzających informacje równolegle i połączonych 187 połączeniami (Akins [1996]) – a żadna reprezentacja nie zawiera „datownika”. Procesy obliczeniowe zachodzące w mózgu czas powstania reprezentacji określają na podstawie jej treści. Łatwiej to zrozumieć na podstawie analogii z przestrzenią – reprezentacje przedmiotów przestrzennych nie muszą znajdować się w konfiguracjach mających kształt tych przedmiotów, tyle że wewnątrz mózgu. Tak samo reprezentacje przedmiotów czasowych nie muszą występować w kolejności odpowiadającej nadchodzeniu bodźców do detektorów układu nerwowego.

W życiu codziennym da się najczęściej orzec, czy dowody przedstawiane w sądzie sfabrykowano przed oskarżeniem kogoś czy też później. Istnieją wyjątki od tej reguły. Jeśli siatka agentów KGB fabrykuje fałszywe dowody przeciwko niewygodnemu działaczowi na rzecz praw człowieka, to część fałszywek może powstać przed kontrolowanym przeciekami do prasy, część później, a część jednocześnie z przeciekami. Załóżmy, że fałszywki te stają się wiarygodne tylko łącznie. Czy można wyróżnić jeden moment, w którym działacz ten stał się obiektem intrygi wywiadowczej? Nie, gdyż fałszywki powstawały niejednocześnie. Słowem – równoległość procesów może uniemożliwiać wyznaczenie precyzyjnego momentu zajścia jakiegoś zdarzenia, gdyż do wyznaczenia tego momentu konieczne jest założenie szeregowego charakteru procesu, w skład którego one wchodzi.

Skoro układ nerwowy ma przetwarzać informacje równolegle, to jak powstaje przeżycie świadome – w szczególności przeżycie odnotowane w opisanych wyżej eksperymentach? Model wielokrotnych szkiców zawdzięcza swoją nazwę metaforze wielu szkicowych wersji artykułu, który krąży wśród specjalistów, poprawiany wielokrotnie przez autora. W pewnym momencie nikt z czytelników nie ma wersji ostatecznej, bo nie wszędzie wszystkie poprawki zostały naniesione; kolejne ulepszenia pojawiają się po drodze. Charakterystyka wszystkich teoretycznych implikacji tego modelu wymagałaby oddzielnego studium; tutaj ograniczę się do minimum (Dennett [1991], s. 111–136):

1. Procesy świadome są realizowane przez równoległe procesy interpretacji i opracowywania informacji, w tym informacji sensorycznych.
2. Elementy reprezentacji świadomych dotyczących obiektów czasoprzestrzennych nie muszą odwzorowywać konfiguracji czasoprzestrzennej tych obiektów (tj. kształtu i kolejności zachodzenia w czasie); elementy te mogą być rozproszone w sieci neuronów.
3. Interpretacja elementów składowych reprezentacji następuje tylko raz; nie następują żadne procesy korekty już utworzonych reprezentacji, są one raczej zastępowane przez nowe reprezentacje.

4. Rozproszone i tworzone równolegle elementy składowe reprezentacji wytwarzają w czasie coś w rodzaju strumienia świadomości, nieustannie korygowanego przez kolejne reprezentacje wytwarzane przez równoległe procesy przetwarzania informacji.
5. W różnych momentach czasowych skupienie uwagi daje różne rezultaty – podmioty wytwarzają różne sprawozdania, które nie mają charakteru niekorygowalnej prawdy na temat procesów świadomych, lecz są raczej próbą spójnej narracji na temat procesów zachodzących w świadomości.
6. Wytwarzanie narracji wytwarza subiektywny ciąg zdarzeń świadomych odczuwany jako jeden strumień świadomości.

Tezy (1)–(6) są empiryczne; odkrycie na przykład ścisłego, geometrycznego odwzorowania obserwowanych obiektów przez pobudzenia neuronów, regularnie zmieniającego się w czasie zgodnie ze zmianami obiektu, oznaczałoby negację tezy (2). Model Dennetta powiada, że strumień świadomości jest wytwarzany przez wiele równoległe działających procesów przetwarzania informacji. Równoległe przetwarzanie ma lepiej tłumaczyć zachodzenie zjawiska ϕ : równoległe procesy interpretują nadchodzące bodźce wzrokowe; efekt gwałtownej zmiany koloru plamki jest wynikiem nałożenia się konkurencyjnych procesów – procesu wykrywającego pojawienie się pierwszej plamki, procesu wykrywającego pojawienie się drugiej plamki, a następnie procesu, który szybko zachodzące zmiany w widzianym obszarze interpretuje jako ruch. To nałożenie można rozumieć jako wielokrotne korekty powstającego szkicu. W odpowiednich warunkach może się okazać, że wynik nałożenia się tych różnych interpretacji tworzy szkic, którego elementy nie są ułożone w kolejności odpowiadającej kolejności zachodzenia postrzeganych zdarzeń (Dennett [1991], s. 136). Nie jest to więc skomplikowany proces tworzenia hipotez ani też korygowania jednolitego strumienia świadomości, lecz kilka równoległych, nakładających się procesów. Postrzeżenia są „datowane” na podstawie treści (złudzenia będą powstawać, gdy na wejściach sensorycznych pojawią się dane wywołujące błędne datowanie na podstawie ich treści). Z tą tezą empiryczną łączy Dennett twierdzenie, że datowanie reprezentacji na podstawie ich treści jest jedynym obiektywnym mechanizmem wyznaczania czasu, jaki jest dostępny podmiotowi.

Z technicznego punktu widzenia kilka równoległe zachodzących procesów obliczeniowych można uznać za jeden, choć skomplikowany proces, przyjmujący wiele różnych danych wejściowych i dający w rezultacie wiele danych na wyjściu (między innymi dlatego na zwykłych, szeregowo pracujących procesorach można wykonywać oprogramowanie zawierające równoległe działające procesy). Względy prostoty i wydajności obliczeniowej mogą przemawiać za tym, że architektura umysłu jest w istocie równoległa. Taką zresztą hipotezę przyjmuje się często w neurokognitywnych teoriach świadomości (por. Baars [2003]). Nie jest to jednak jedyna możliwa poprawna hipoteza tłumacząca pojawianie się zja-

wiska $\varnothing$. Mamy tutaj więc do czynienia z pewnymi wystarczającymi warunkami zachodzenia tego zjawiska; nie ma jednak mowy o tym, że są to warunki konieczne⁹. Nie wiadomo, czy są one też faktyczne.

4. Wysoka cena inżynieryjnych wyjaśnień

Zadaniem inżynierii odwrotnej jest dostarczanie hipotez na temat mechanizmów – czy to biologicznych, czy to psychologicznych. Mechanizmy te mają zaś tłumaczyć, czyli wyjaśniać działanie złożonych układów takich jak umysł. Jeśli stanowisko naturalistyczne w filozofii ma nie ograniczać się tylko do formalnej analizy spójności ewentualnego języka opisującego tradycyjnie kłopotliwe dla naturalistów zjawiska – przede wszystkim umysł – to musi ono sięgać po wyjaśnienia. Nie mogą one być jednak natury pojęciowej. Muszą to być wyjaśnienia w sensie właściwym – wyjaśnienia zapewniane przez teorie empiryczne. Jeśli zaś nie ma jeszcze odpowiednich teorii, to należy odwołać się przynajmniej do prototeorii; w przeciwnym wypadku, jeśli pozostanie coś, czego *nie można* wyjaśnić zgodnie z naukami przyrodniczymi, wówczas teza naturalizmu eksplanacyjnego okaże się fałszywa.

Inżynieryjna analiza służy więc nie tylko wyjaśnianiu mechanizmów umysłu. Nadaje ona wyższe prawdopodobieństwo tezie naturalizmu eksplanacyjnego. Naturalizm eksplanacyjny z kolei ma ambicję stworzenia szerokiego, racjonalnego obrazu świata we współpracy z nauką. Naturalistyczne rozwiązanie problemu psychofizycznego nie może pozostać jedynie konceptualno-terminologiczne, lecz musi zawierać też wyjaśnienie znanych nam cech umysłu za pomocą jego funkcjonalnych modułów. Rzecz jasna, będzie to rodzaj wyjaśnienia złożonych obiektów przez ich części składowe oraz relacje między nimi; słowem – rodzaj wyjaśnienia redukcyjnego. Uzyskane wyjaśnienia – usprawiedliwiające naturalizm eksplanacyjny – stają się elementem filozoficznie uzasadnianej wizji świata. Filozofia nie jest izolowana od empirii.

Takie stanowisko jest narażone na wiele niebezpieczeństw. Po pierwsze, hipotezy naukowe starzeją się zwykle szybciej od filozofii, szukającej najczęściej apriorycznych prawd. Jednak odcinając się od hipotez naukowych, filozof wystawia się na ryzyko gabinetowej spekulacji lub tworzenia pozbawionych implikacji empirycznych systemów dedukcyjnych. Po drugie, hipotezy te nigdy nie mogą rościć sobie prawa do wyłączności – z natury rzeczy mogą określić tylko wystarczające warunki w wyjaśnieniach. Filozofia zaś najczęściej miała dostarczać prawd koniecznych. Po trzecie, wymaga ono zaangażowania się po stronie redukcjonizmu i mechanicyzmu w wersji funkcjonalistycznej. Te stanowiska bywają współ-

⁹ Istnieje wiele innych hipotez, które mają tłumaczyć pojawianie się tych zjawisk, np. w polskiej literaturze przedmiotu Piłat [1999], s. 145.

częście krytykowane, zwłaszcza przez antyredukjonistów (por. np. Kim [2002]). Filozofia umysłu korzystająca przede wszystkim z inżynierii odwrotnej musi się odwołać do zawodnych, empirycznych hipotez, zrezygnować z roszczenia do prawdy koniecznej na temat umysłu, a wreszcie opowiedzieć się za redukcjonizmem. Myśliciele naturalistyczni gotowi są skądinąd do odrzucenia ambicji tradycyjnej filozofii, gdyż za tę cenę mogą skorzystać z wiedzy naukowej, która jako jedyna może być uznawana za wiedzę w pełnym sensie tego słowa. Jednak równie uprawnionym stanowiskiem jest rezygnacja z uroszczeń filozofii do rozstrzygania sporów empirycznych i uznanie, że filozof po prostu nie powinien oferować żadnych teorii wyjaśniających, o ile ma pozostać filozofem. Może je zająć chociażby zwoleńnik tradycyjnej analizy pojęciowej.

Stosowane przez Dennetta metody pozwalają mu rozstrzygać więcej problemów niż wielu innym filozofom analitycznym. Filozofowie języka potocznego nie są w stanie stawiać problemów kognitywnych, a więc oferować własnej teorii wyjaśniającej umysł. Wielbiciele eksperymentów myślowych, najczęściej odwołujących się do złożonej aparatury logik modalnych, nie są też w stanie wykroczyć poza granice analizy języka potocznego, aby podjąć problemy kognitywne. Naturaliści ontologiczni muszą ograniczać się do opisu czy analiz pojęciowych. Z kolei fenomenologowie często odcinają się od naukowych źródeł poznania, co jest szczególnie widoczne w postmodernistycznych kontynuacjach fenomenologii. Ci, którzy korzystają z dokonań nauki, posługują się wówczas metodami innymi niż fenomenologia, gdyż ta nie służy przecież do tworzenia czy bezpośredniej oceny teorii empirycznych, lecz raczej do opisu pewnych istotnych, pojęciowych czy intuicyjnych, zależności. Dlatego też ambicją fenomenologii nie jest i nie może być wyjaśnianie zjawisk empirycznych. Filozofowie konstruujący ontologie formalne – ale nie rekonstruujący ich z wyników nauki – nie mogą, tak samo jak ontologowie naturalistyczni, dostarczyć żadnych wyjaśnień. Na tym tle naturalizm eksplanacyjny, zapewniający filozofowi analitycznemu możliwość tworzenia szerokiego, spójnego i racjonalnego obrazu świata, wydaje się kierunkiem obiecującym, o ile tylko przyjmie się założenie, że cena, jaką za to przyjdzie zapłacić, nie jest zbyt wygórowana. Choć szczegółowe rozwiązania problemów kognitywnych u Dennetta mogą się szybko zestarzeć, to sama idea tego rodzaju filozofii ma szansę na przetrwanie. Umożliwia ona bowiem godzenie dobrze uzasadnionych intuicji zdrowego rozsądku, w tym intuicji dotyczących umysłu, z naturalistyczną wizją świata. Dzięki temu konstrukcja umysłu, którą odtwarza Dennett, jest zarówno zgodna z podstawowymi przeświadczeniami zdrowego rozsądku – takimi jak przekonanie, że przeżycia świadome są rzeczywiste i że opisy intencjonalne są praktycznie niezbędne – jak i z poważnie traktowanymi wynikami nauki.

Bibliografia

- Akins, K., 1996, *Lost the Plot? Reconstructing Dennett's Multiple Drafts Theory of Consciousness*, "Mind & Language", 11, 1/1996, s. 1–46.
- Allen, C. i Bekoff, M., 1995, *Biological Function, Adaptation, and Natural Design*, "Philosophy of Science", 62, nr 4, s. 609–622.
- Amundson, R. i Lauder, G., 1998, *Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology*, w: Hull i Ruse [1998], s. 227–257.
- Baars, B., 2003, *How Brain Reveals Mind. Neural Studies Support the Fundamental Role of Conscious Experience*, "Journal of Consciousness Studies", 10, nr 9–10, s. 100–114.
- Brockman, J. (red.), 1996, *Trzecia kultura*, tł. różni, Wydawnictwo CiS, Warszawa.
- Csete, M.E., Doyle, J.C., *Reverse Engineering of Biological Complexity*, "Science", 3/01/2002, 295, nr 5560, s. 1664–1670.
- Cummins, R., 1975, *Functional Analysis*, "The Journal of Philosophy", 72, nr 20, s. 741–765.
- Cutland, N., 1980, *Computability*, Cambridge University Press, Cambridge.
- Dennett, D. i Kinsbourne, M., 1992, *Time and the Observer: the Where and When of Consciousness in the Brain*, "Behavioral and Brain Sciences", 15, s. 183–247.
- Dennett, D., 1991, *Consciousness Explained*, Little, Brown, Boston.
- Dennett, D., 1995, *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, Simon & Schuster, Nowy Jork.
- Dennett, D., 1996, *Dźwięnie wyobraźni*, przeł. W. Turopolski, w: Brockman [1996], Wydawnictwo CiS, Warszawa.
- Dennett, D., 1998, *Brainchildren: Essays on Designing Minds*, MIT Press, Cambridge (Mass.).
- Dennett, D., 2003a, *Freedom Evolves*, Viking.
- Dennett, D., 2003b, *What RoboMary Knows*, w: *Knowledge Argument*, Torin Alter (red.) (wersja robocza dostępna pod adresem <http://ase.tufts.edu/cogstud/papers/RoboMaryfinal.htm>).
- Dennett, D., 2003c, „Naprawdę przekonani: strategia intencjonalna i dlaczego ona działa”, przeł. M. Miłkowski, *Przegląd Filozoficzno-Literacki* 4(6)/2003, s. 87–109.
- Godfrey-Smith, P., 1998a, *Maternal Effects: On Dennett and Darwin's Dangerous Idea*, "Philosophy of Science", 65, s. 709–720.
- Goodman, N., 1997, *Jak tworzymy świat*, przeł. M. Szczubiałka, Fundacja Aletheia, Warszawa.
- Griffiths, P.E., 1996, *The Historical Turn in the Study of Adaptation*, "The British Journal for the Philosophy of Science", 47, nr 4, s. 511–532.

- Hull, D. i Ruse, M., (red.), 1998, *The Philosophy of Biology*, Oxford University Press.
- Kim, J., 2002, *Umysł w świecie fizycznym*, przeł. R. Poczobut, Wydawnictwo IFiS PAN, Warszawa.
- Klawiter, A., 1999, *Wprowadzenie. O naturze problemu kognitywnego*, w: *Poznańskie Studia z Filozofii Humanistyki. Tom 5 (18). Umysł a rzeczywistość*, A. Klawiter, L. Nowak, P. Przybysz (red.), Zysk i S-ka. Wydawnictwo, Poznań, s. 135–138.
- Lewens, T., 2002, *Adaptationism and Engineering*, "Biology & Philosophy", 17, s. 1–31.
- Maynard Smith, J., 1995, *Genes, Memes & Minds*, "The New York Review of Books", 42, nr 19, 30 listopada 1995.
- Millikan, R.G., 1984, *Language, Thought, and Other Biological Categories. New Foundations for Realism*, MIT Press, Cambridge (Mass.).
- Piłat, R., 1999, *Umysł jako model świata*, IFiS PAN, Warszawa.
- Pinker, S., 2002, *Jak działa umysł*, przeł. M. Koraszewska, Książka i Wiedza, Warszawa.
- Richardson, R.C., 2003, *Engineering Design and Adaptation*, "Philosophy of Science", 70, s. 1277–1288.
- Wright, L., 1973, *Functions*, "The Philosophical Review", 82, nr 2, s. 139/168.

Philosophy as Reverse Engineering: The Case of Daniel C. Dennett

Daniel C. Dennett approaches philosophy of mind in a manner typical for reverse engineering. He studies the design and behaviour of a mechanism in order to discover what the functions of various units in the system could be. He applies this method to explain some biological facts and then to interpret specifically mental operations. The author tries to establish why Dennett uses this approach and proposes that his main reason may have been to sustain a naturalistic conception of mind.