

# Daniel C. Dennett

OD BAKTERII  
DO BACHA  
O EWOLUCJI UMYSŁÓW

Tłumaczenie

Krystyna Bielecka i Marcin Miłkowski



## 4. Dwa osobliwe odwrócenia rozumowania

### Jak Darwin i Turing dokonali odczarowania

Świat przed Darwinem uspołniała nie nauka, lecz tradycja. Wszystkie rzeczy we wszechświecie, od istoty najbardziej wywyższonej („człowiek”) do najbardziej przyziemnej (mrówka, żwir, kropelka deszczu), były dziełami jeszcze wyższej istoty, Boga, wszechmogącego i wszechwiedzącego inteligentnego twórcy – który cechował się uderzającym podobieństwem do drugiej co do wywyższenia istoty. Nazwijmy to „teorią stworzenia jako skapywania”. Darwin zastąpił ją teorią stworzenia jako kipienia. Robert MacKenzie Beverley<sup>12</sup>, jeden z dziewiętnastowiecznych krytyków Darwina, ujął to obrazowo:

W teorii, z którą mamy do czynienia, absolutna ignorancja jest wynalazcą; moglibyśmy więc sformułować podstawową zasadę całego systemu: **ABY ZBUDOWAĆ DOSKONAŁĄ I PIĘKNĄ MASZYNĘ, NIE TRZEBA WIEDZIEĆ, JAK JĄ ZBUDOWAĆ.**  
Po starannej analizie okaże się, że teza ta w postaci skróconej

---

<sup>12</sup> Od ponad trzydziestu lat błędnie określałem tego autora jako Roberta Beverleya MacKenziego; dziękuję kompetentnym redaktorom wydawnictwa Norton za wprowadzenie mnie z błędu.

wyraża zasadniczy cel teorii, a także w kilku słowach wszystko, co pan Darwin miał na myśli; ów, osobiwie odwracając rozumowanie, zdaje się uważać, iż absolutna ignorancja mogłaby zastąpić absolutną mądrość we wszystkich twórczych osiągnięciach. (Beverley 1868)

To było rzeczywiście „osobliwe odwrócenie rozumowania”, a niedowierzanie wyrażane przez Beverleya nadal pobrzmiewa u zniechęcająco dużego odsetka ludności w XXI wieku. Kiedy przejdziemy do Darwinowskiej teorii stworzenia jako kipienia, możemy metaforycznie pojmować wszystkie twórcze prace projektowe jako podnoszenie w tak zwanej przeze mnie przestrzeni projektowej. Musi zacząć się od pierwszych, prymitywnych replikatorów, jak widzieliśmy w rozdziale 3, i stopniowo wznosić się, fala po fali doboru naturalnego, aż do życia wielokomórkowego we wszystkich jego formach. Czy taki proces rzeczywiście potrafił stworzyć wszystkie dziwy obserwowane w biosferze? Sceptycy już od czasów Darwina starali się udowodnić, że taki czy inny fenomen po prostu nie może się pojawić na takiej pracochłonnej i nieinteligentnej drodze. Szukali czegoś żywego, ale *niemogącego wyewoluować*. Takie zjawisko określam terminem „latający dywan” (*skyhook*) na cześć mitycznych środków transportu, za pomocą których można unosić się w górę (Dennett 1995a). Latający dywan unosi się wysoko w przestrzeni projektowej, niewspierany przez przodków, będąc bezpośrednim wynikiem szczególnego aktu inteligentnego tworzenia. Owi sceptycy nie odkryli ani razu cudownego latającego dywanu, ale raz za razem tylko wspianiałego żurawia, niecudowny wynalazek w przestrzeni projektowej, który pozwala na jeszcze efektywniejsze badanie możliwości projektowania, jeszcze wydajniej-

sze podnoszenie w przestrzeni projektowej. Endosymbioza jest żurawiem; podniosła proste pojedyncze komórki w sferę znacznej złożoności, gdzie mogło zacząć się życie wielokomórkowe. Płeć jest żurawiem; pozwoliła na mieszanie się puli genetycznej, a tym samym na o wiele bardziej skuteczne próbkowanie przez ślepe procesy prób i błędów doboru naturalnego. Język i kultura to żurawie, nowości ewolucyjne, które otworzyły rozległe przestrzenie możliwości do zbadania przez coraz bardziej inteligentnych (ale nie cudownie inteligentnych) projektantów. Bez dodatku języka i kultury do arsenału narzędzi badawczo-rozwojowych dostępnych ewolucji nie istniałyby rośliny tytoniu świecące w ciemności z genami robaczek świętojańskich. One nie są cudowne. Są one tak samo ewidentnie owocami drzewa życia, jak sieci pajaków i tamy bobrów, ale prawdopodobieństwo ich pojawienia się bez pomocy *Homo sapiens* i naszych narzędzi kulturowych jest zerowe.

Kiedy coraz lepiej poznajemy nanomaszynę życia, która za to wszystko odpowiada, możemy docenić drugie osobliwe odwrócenie rozumowania, przeprowadzone prawie sto lat później przez innego genialnego Anglika: Alana Turinga. Oto osobliwe odwrócenie Turinga, wypowiedziane w języku Beverleya:

ABY BYĆ DOSKONAŁĄ I PIĘKNĄ MASZYNĄ OBLICZENIOWĄ, NIE TRZEBA WIEDZIEĆ, CZYM JEST ARYTMETYKA.

Przed wynalazkiem Turinga istniały setki lub tysiące rachmistrzów (*computers*), pracujących nad obliczeniami naukowymi i inżynierskimi. Rachmistrzowie to ludzie, a nie maszyny. Wielu z nich to kobiety, często wykształcone matematycznie na

uczelniah wyższych. To byli ludzie, którzy wiedzieli, co to jest arytmetyka, ale Turing pojął coś bardzo ważnego: nie muszą tego wiedzieć! Jak zauważył, „zachowanie komputera w danej chwili zależy od symboli, które obserwuje, a także jego »stanu umysłu« w tej chwili” (Turing 1936, s. 5). Ten „stan umysłu” (w cudzysłowie u Turinga) był prościutkim zestawem instrukcji „jeśli – to” dotyczących tego, co należy zrobić i do jakiego „stanu umysłu” przejść (i powtarzać, dopóki nie pojawi się instrukcja STOP). Turing pokazał, że można zaprojektować bezmyślne maszyny, które były absolutnie ignoranckie, ale które mogły idealnie wykonywać arytmetykę, postępując wedle „instrukcji” zrealizowanych mechanicznie. Co ważniejsze, wykazał, że jeśli ich instrukcje zawierają *skoki warunkowe* (instrukcje „jeśli – to”, takie jak „jeśli obserwujesz liczbę 0, zastąp ją liczbą 1 i przesun się w lewo, a jeśli zauważysz 1, zostaw to, co jest, i przesun się w prawo, a także zmień stan na  $n$ ”), to te maszyny mogłyby prowadzić nieskończenie skomplikowane ścieżki określone przez instrukcje, co zapewniało im niezwykle umiejętności: mogłyby obliczyć *wszystko*, co tylko daje się obliczyć. Innymi słowy, programowalny cyfrowy komputer to uniwersalna maszyna Turinga, zdolna do naśladowania dowolnego cyfrowego komputera specjalnego przeznaczenia, wykonująca zestaw instrukcji implementujących programowo komputer specjalnego przeznaczenia<sup>13</sup>. (Nie trzeba zmieniać elektroniki

---

<sup>13</sup> Standardowy żargon służący do wyrażenia tej myśli jest znany jako teza Churcha i Turinga, sformułowana przez logika Alonza Churcha: „wszystkie procedury efektywne są obliczalne w sensie Turinga” – choć oczywiście wiele z nich nie jest w praktyce obliczalnych, gdyż ich wykonanie zajmuje za dużo czasu. Ponieważ nasze rozumienie, które procedury (czyli w gruncie rzeczy programy komputerowe lub algorytmy) są efektywne, jest w sposób nieunikniony intuicyjne, ta teza nie może zostać udowodniona, ale jest niemal powszechnie przyjmowana, więc obliczalność w sensie Turinga jest zazwyczaj uznawana za dopuszczalną definicję operacyjną efektywności.

w smartfonie, aby wykonał on nowe zadania, wystarczy pobrać aplikację i przekształcić go w wyszukiwarkę gwiazd, translator, kalkulator lub korektor pisowni...) Ogromna przestrzeń projektowa przetwarzania informacji została otwarta przez Turinga, który przewidział, że istnieje długa ścieżka wiodąca od całkowitej ignorancji do sztucznej inteligencji, przez długi szereg kroków w górę w tej przestrzeni projektowej.

Wielu ludzi nie może znieść Darwinowskiego osobliwego odwrócenia. Nazywamy ich „kreacjonistami”. Nadal poszukują latających dywanów – „nieredukowalnie złożonych” (Behe 1996) cech biosfery, które nie mogły powstać w wyniku ewolucji przebiegającej w procesach darwinowskich. Wielu ludzi nie może znieść Turingowskiego osobliwego odwrócenia, także z uderzająco podobnych powodów. Chcą wierzyć, że cuda umysłu nie mogą być wytworzone przez zwykłe procesy materialne, że umysły są tajemnicze, jeśli nie dosłownie cudowne, gdyż wymykają się naukom przyrodniczym. Nie chcą, aby rana kartezyńska została zaleczona.

Dlaczego nie? Opisaliliśmy już niektóre z ich mniej chwalebnych motywów: strach, dumę, źle ulokowaną miłość do nierozwiązanej tajemnicy. Oto inny powód (czy to z *jakiej przyczyny*, czy *po co?*): zarówno Darwin, jak i Turing twierdzą, iż odkryli coś naprawdę niepokojącego ludzki umysł – *że można umieć, ale nie rozumieć*. Beverley wyraził swoje oburzenie z entuzjazmem: *sama idea* istnienia umiejętności twórczych bez inteligencji! Zastanówmy się, jak głęboko przeciwstawia się to myśli zapisanej w naszej polityce i praktyce edukacyjnej: (*najlepiej*) *rozumieć, aby coś umieć*. Wysyłamy dzieci na uniwersytety, aby pod każdym względem zrozumiały, jak funkcjonuje świat, bo przyda im się to przez całe ich życie, zapewniając

potrzebne umiejętności pochodzące z wartościowego rozumienia, które im wpojono. (Nawiasem mówiąc, używam terminów „rozumieć” [*comprehension*] i „rozumienie” [*understanding*] zamiennie, chociaż forma czasownikowa pojawia się częściej ze względu na stosowany przeze mnie środek stylistyczny.) Dlaczego dziś lekceważymy uczenie się na pamięć? Ponieważ odkryliśmy – nieprawdaż? – że jeśli (tylko jeśli lub może najlepiej jeśli?) dzieci *rozumieją* zagadnienie bądź metodę, to umieją sobie z nimi radzić. Lekceważymy tępe powtarzanie, polegające na wypełnianiu szablonu bez zrozumienia, o co w nim chodzi. Szydzimy z założenia, że kolorowanki mogłyby służyć do wyszkolenia twórców. Nasze motto mogłoby równie dobrze brzmieć:

Jeśli nauczysz je rozumieć, będą także umieć!

Zauważmy, że w grę wchodzi tu coś więcej niż tylko ciut ideologii. Znamy pewne katastrofalne, błędne zastosowania naszej świętej zasady, na przykład „nową matematykę”, w ramach której starano się – bez powodzenia – najpierw nauczyć dzieci teorii mnogości i innych abstrakcyjnych pojęć zamiast ćwiczenia ich w dodawaniu i odejmowaniu, tabliczce mnożenia, ułamkach i prostych algorytmach, takich jak dzielenie liczb wielocyfrowych czy liczenie dwójkami, piątkami i dziesiątkami.

Siły zbrojne to jedno z najbardziej skutecznych instytucji oświatowych na świecie. Przemieniają przeciętnych uczniów szkół średnich w niezawodnych mechaników silników odrzutowych, operatorów radarowych, nawigatorów i wielu innych specjalistów technicznych dzięki dużym dawkom „ćwiczeń i utrwalania”. W odpowiednim czasie u tych wyćwiczonych praktyków z zaszczepionych umiejętności powstaje wiele wartościowego

rozumienia, więc mamy dobre empiryczne dowody, że umiejętności nie zawsze zależą od rozumienia, a czasami są wręcz warunkiem wstępnym rozumienia. Darwin i Turing przedstawili najskrajniejszą wersję tej obserwacji: *wszelki* talent i zrozumienie na świecie wyrastają w ostatecznym rozrachunku z bezrozumnych umiejętności łączonych z czasem w coraz więcej umiejące, a więc i rozumiejące, systemy. Jest to rzeczywiście osobliwe odwrócenie, obalenie dającej prymat umysłowi, przeddarwinowskiej wizji stworzenia przez wizję ewolucji nas, inteligentnych projektantów, a zatem przez wizję spychającą umysł na *sam koniec*.

Nasz sceptycyzm co do tego, czy można umieć, ale nie rozumieć, ma pewne przyczyny, lecz nie racje za sobą. Nie jest „racjonalne” przekonanie, że nie można umieć, ale nie rozumieć; to po prostu wydaje się słuszne, a wydaje się słuszne, ponieważ nasze umysły zostały ukształtowane, by myślały w ten sposób. Dopiero Darwin odczarował tego rodzaju magiczne myślenie, a Turing wkrótce potem uczynił to ponownie, proponując, że można odwrócić tradycyjny porządek i budować rozumienie z kaskady umiejętności w z grubsza taki sposób, jak ewolucja poprzez dobór naturalny wytwarza coraz bardziej olśniewające układy wewnętrzne, narządy i instynkty bez rozumienia, co robią.

Jest duża różnica między osobliwym odwróceniem u Darwina i u Turinga. Darwin pokazał, jak genialne projekty mogłyby powstać dzięki kaskadom procesów pozbawionych w ogóle inteligencji, ale system kaskad procesów Turinga był wytworem bardzo inteligentnego projektanta, Turinga. Można by powiedzieć, że chociaż Darwin odkrył *ewolucję* przez dobór naturalny, Turing *wynalazł* komputer. Wielu twierdzi, że inteligentny Bóg musiał ustalić wszystkie warunki ewolucji na drodze



doboru naturalnego, a Turing odgrywa bodaj tę właśnie rolę, tworząc podstawową ideę komputera (materialnego, nieożywionego, nierozumiejącego), mogącego stać się areną, na której rozumienie może wyrastać z czegoś podobnego do ewolucji, szeregu udoskonaleń konstrukcyjnych, wymyślonych z podstawowych cegiełek obliczeniowych. Czyż rola Turinga jako inteligentnego projektanta nie *przeciwstawia się* osobliwemu odwróceniu Darwina, a nie stanowi jego *kontynuacji*? Nie, a odpowiedź na to ważne pytanie jest głównym zadaniem dalszych części tej książki. Krótkie wyjaśnienie jest następujące: sam Turing to jedna z gałązek drzewa życia, a jego artefakty, konkretne i abstrakcyjne, są pośrednio wytworami ślepych procesów darwinowskich, tak jak sieci pajaków i tamy bobrów, więc nie ma *radykałnej nieciągłości*, nie trzeba latającego dywanu, aby od pajaków i bobrów dotrzeć do Turinga i jego maszyn. Mimo to istnieje duża luka, którą należy wypełnić, ponieważ to, co Turing robił, było uderzająco odmienne od metod pajaków i bobrów. Potrzebujemy dobrego ewolucyjnego wyjaśnienia tej różnicy. Jeśli umiejętności *bez* zrozumienia są tak cudownie płodne – mogą w końcu zaprojektować słowiki – to dlaczego potrzebujemy rozumienia – zdolnego do projektowania ód na cześć słowików i komputerów? Dlaczego i jak ludzkie rozumienie w ogóle się pojawiło? Najpierw wyraźniej i ostrzej zarysujmy ten kontrast.

Jeśli termyty to imponujące przykłady umiejętności bez rozumienia, gdyż potrafią budować mocne, bezpieczne, klimatyzowane domy bez planów ani szefów (królowa termitów przypomina bardziej klejnoty koronne niż szefa), to Antoni Gaudí jest niemal doskonałym modelem inteligentnego projektanta, boskiego szefa, uzbrojonego od samego początku w rysunki, plany

i orędzia pełne wyłożonych z werwą racji. Jego wielki kościół w Barcelonie to wyjątkowy przykład tworzenia odgórnego, ale oryginalny komputer Turinga, Pilot ACE (który można zobaczyć w Muzeum Nauki w Londynie), może pokonać go w konkurencji o pierwszą nagrodę. Jeden z pierwszych naprawdę pożytecznych komputerów został uruchomiony w 1950 roku w National Physical Laboratory w Anglii i rywalizował z Sagrada Família pod względem oryginalności, złożoności – i kosztów. Obaj twórcy musieli przekonywać zwolenników do finansowania ich ambitnych projektów, a także opracowali skomplikowane schematy wraz z wyjaśnieniami. W obu przypadkach zaistnienie wytworu zależało od istnienia wcześniejszych reprezentacji, w umyśle geniusza, celu projektu, a tym samym *raison d'être* wszystkich jego części<sup>14</sup>. Gdy przyszło do rzeczywistej budowy artefaktów, w grę wchodziło pracownicy, którzy *stosunkowo* mało rozumieli, minimalnie doceniali cel swych prac. Rozumienie było oczywiście rozproszone: Gaudí nie musiał rozumieć dokładnie, jak mieszać zaprawę lub rzeźbić kamienie tak jak robotnicy czy kamieniarze, a Turing nie musiał być wirtuozem lutownicy ani ekspertem w dziedzinie produkcji lamp elektronowych. Tego rodzaju rozproszenie wiedzy fachowej czy rozumienia jest znakiem rozpoznawczym ludzkich projektów twórczych i jest zdecydowanie kluczowe dla dzisiejszych zaawansowanych technicznie wyrobów, ale nie dla wszystkich wcześniejszych artefaktów. Można samodzielnie zrobić włócznię, a nawet kajak, drewniany wóz lub chatę ze strzechą, rozumiejąc

---

<sup>14</sup> Gaudí zmarł w 1926 roku, ale pozostawił rysunki, instrukcje i modele, które wciąż kierują pracami zmierzającymi do zbudowania niedokończonego kościoła; Turing porzucił NPL przed zakończeniem prac nad Pilot ACE, ale zostawił również reprezentacje artefaktu mające służyć do jego ukończenia.

każdy aspekt projektu i budowy, ale już nie radio, samochód czy elektrownię jądrową.

Bliższa analiza kilku przykładów ludzkich wyrobów i wynalazków technicznych stworzonych do ich produkcji posłuży do wyjaśnienia stacji pośrednich na drodze od bezmyślnych bakterii do Bacha, ale najpierw musimy wprowadzić termin, który pojawił się w filozofii, a następnie jest wykorzystywany w kilku przedsięwzięciach naukowych i inżynierskich.