Filozofia Nauki (*The Philosophy of Science*)
ISSN 1230-6894 e-ISSN 2657-5868
2020, vol. 28(4) [112]: 83-102
DOI: 10.14394/filnau.2020.0023

MACIEJ TARNOWSKI*

JAKA TEORIA DZIAŁANIA?
O *MECHANICE DZIAŁAŃ* MICHAŁA BARCZA

Michał Barcz, *Mechanika działań. Filozoficzny spór wokół przyczynowej teorii działania*, Warszawa: Wydawnictwo Naukowe PWN 2020, ss. 267, ISBN 978-83—01-21323-7

Abstract

WHAT THEORY OF ACTION? ON MICHAŁ BARCZ'S *MECHANICS OF ACTIONS*

This is a review of Michał Barcz's book *Mechanika działań. Filozoficzny spór wokół przyczynowej teorii działania* (*Mechanics of Actions: Philosophical Dispute over the Causal Theory of Action*). The book discusses various causal accounts of intentional action and presents several arguments against them. The review focuses on accuracy in presentation of different causal theories of action and soundness of arguments presented against them by the author. At the end, I discuss some methodological issues raised by the book.

Keywords: philosophy of action, causal theory of action, intention, rationality, artificial intelligence

Chociaż filozofia działania przynajmniej od lat 80. XX wieku stanowi samodzielną i prężnie rozwijającą się dziedzinę filozofii analitycznej, dopiero teraz doczekaliśmy się pierwszej dłuższej rozprawy w języku polskim poświęconej szczegółowo jej zagadnieniom. Jest to o tyle zaskakujące, że jest to dziedzina istotna dla wielu nauk — rozstrzygnięcia formułowane w niej mają wpływ zarówno na prawo i etykę, jak i na rozwój kognitywistyki, sztucznej inteligencji czy robotyki. Ponadto nieformalny ojciec tej dyscypliny, Donald Davidson, cieszył się zawsze w Polsce dużą estymą i zainteresowaniem. Symptomatyczne jest

* Wydział Filozofii, Uniwersytet Warszawski, ul. Krakowskie Przedmieście 3, 00-927 Warszawa, e-mail: m.tarnowski3@student.uw.edu.pl, ORCID: https://orcid.org/0000-0003-3824-4134.

jednak, że w najważniejszym dla polskiej recepcji jego filozofii wyborze *Eseje o prawdzie, języku i umyśle* pod redakcją Barbary Stanosz brak właśnie tekstów Davidsona stanowiących podwaliny jego filozofii działania.

Należy więc stwierdzić, że książka Michała Barcza *Mechanika działań. Filozoficzny spór wokół przyczynowej teorii działania* wypełnia niezwykle istotną lukę na polskojęzycznym rynku wydawniczym. Jak na standardy owego rynku jest to również książka bardzo dobrze wydana: czytelny skład i spis treści, bibliografia oraz indeks pojęć i osób pozwalają dobrze odnaleźć się podczas lektury. Należy także pochwalić korektę — błędy są niezwykle rzadkie, natomiast samo wydanie jest bardzo estetyczne. Wszystkie te cechy sprawiają, że czytelnicy bez przeszkód mogą korzystać z książki, która w wielu wypadkach będzie ich pierwszym kontaktem z poruszaną tematyką.

Dlatego należy zastanowić się, czy książka Barcza pozwoli dotrzeć do bogactwa filozofii działania laikom (taką opinię można znaleźć w cytowanej na tylnej okładce recenzji Roberta Piłata) i czy faktycznie taki cel Barcz sobie stawia. Tematem książki jest przyczynowa teoria działania — koncepcja, zgodnie z którą „tym, co stanowi naturę działań, są ich umysłowe przyczyny (stany umysłu) będące jednocześnie przemawiającymi za nimi racjami" (Barcz 2020: 14). Ponieważ nadal pozostaje ona poglądem dominującym i w kontrze do niej powstawały alternatywne poglądy na naturę działań, zapoznanie się z nią oraz jej problemami pozwoli czytelnikom zyskać dobrą orientację we współczesnej debacie. Książka Barcza może więc posłużyć jako wstęp do całej dziedziny, tym bardziej że nie wymaga uprzedniego zaznajomienia czytelników z filozofią analityczną.

Ambicje autora wykraczają jednak poza zwyczajne przedstawienie teorii przyczynowej — Barcz, jak to zaznacza we wprowadzeniu, zamierza przekonać czytelnika, że teoria przyczynowa jest „dalece niesatysfakcjonująca" (Barcz 2020: 15), oraz wysunąć wobec niej oryginalne zarzuty. Jak widać, dzieło Barcza może więc zrealizować trzy zasadnicze cele: (a) pedagogiczny i popularyzatorski (przedstawienie podstawowych ujęć i problemów teorii przyczynowej), (b) krytyczny (sformułowanie oryginalnych zarzutów) oraz (c) metodologiczny (omówienie statusu teorii przyczynowej we współczesnej filozofii działania). Rekonstrukcję różnych ujęć teorii przyczynowej — w szczególności koncepcji Davidsona, Alfreda Melego i Freda Dretskego — znajdziemy w pierwszej części książki (rozdziały 1-4), natomiast omówienie klasycznych zarzutów przeciwko niej (argumenty ze związku logicznego, dziwnych ciągów przyczynowych i zagubionego sprawcy) w rozdziałach piątym, szóstym i siódmym. Treść tych rozdziałów omówię więc przede wszystkim z perspektywy trafności merytorycznej i przystępności zawartego w nich opisu. Rozdziały ósmy i dziewiąty przedstawiać mają zgodnie z zapowiedzią autora jego oryginalne kontrargumenty,

obejmujące problem z teoretycznym ujęciem działań omyłkowych oraz zgodność z paradygmatem koneksjonistycznym w badaniach nad sztuczną inteligencją. Te dwa rozdziały omówię krytycznie, polemizując z niektórymi ustaleniami Barcza. Pod koniec postaram się podjąć metodologiczną refleksję podniesioną w książce: odpowiem na pytanie, czy teoria przyczynowa jest wciąż owocnym paradygmatem w analizie działania i jaką ma konkurencję.

1. REKONSTRUKCJA PRZYCZYNOWEJ TEORII DZIAŁAŃ

Pierwszą część książki otwiera obszerne omówienie krytyczne teorii Davidsona (rozdz. 1 i 2). Ponieważ filozof ten nie przedstawił nigdy poglądów na naturę działania w ujednoliconej formie (na jego dorobek w tej dziedzinie składają się zarówno eseje pisane w ciągu bez mała trzydziestu lat od publikacji *Actions, Reasons and Causes*, jak i rozmaite tropy pozostawione w tekstach poświęconych teorii znaczenia czy ontologii umysłu), ich jasna i niekontrowersyjna rekonstrukcja jest zadaniem niezwykle trudnym. Cieszy więc, że autor *Mechaniki działań* wywiązuje się z tego zadania bardzo dobrze i zwięźle. Co szczególnie istotne, w klarowny sposób zostają tutaj omówione i wytłumaczone pojęcia związane z Davidsonowską teorią interpretacji radykalnej („interpretacja", „zasada życzliwości" czy założenie o racjonalności opisywanego podmiotu) oraz argumentacja za stanowiskiem monizmu anomalnego w filozofii umysłu. Jedyną trudnością dla laika może być fakt dość losowego stosowania wytłuszczeń (np. słowa „natura" na s. 14, „nie zależy" na s. 36 czy „jedynie przypadkiem" na s. 64), które nie wskazują (wbrew oczekiwaniom) podstawowych pojęć, a jedynie akcentują wypowiedź autora. Być może rozsądniejsze z perspektywy niewprowadzonego czytelnika byłoby wytłuszczenie problematycznych pojęć i zamieszczenie ich tłumaczenia lub słowniczka.

Mankamentem tego rozdziału jest jego niekonkluzywność: zarzut epifenomenalizmu treści wysunięty przez Kima (1989) wobec monizmu anomalnego nie został odniesiony do Davidsonowskiej teorii działania. Nie sposób tego zrobić samodzielnie, gdyż w rozdziale brakuje omówienia związków logicznych między tymi koncepcjami. W rozdziale drugim, zawierającym rekonstrukcję krytyki sformułowanej przez Louise Anthony (1989), Barcz *de facto* uznaje, że koncepcje te są sprzeczne — być może jednak wciąż możliwe jest, że ontologiczna podbudowa koncepcji Davidsona (monizm anomalny) jest niezależna logicznie od jego teorii działania. Wówczas nie możemy, wbrew intencji autora, przyjąć negatywnego poglądu na temat samej teorii przyczynowej w wersji Davidsona. Przedstawiony atak na koncepcję monizmu anomalnego

wydaje się teoretycznie niezależny od słuszności poglądów Davidsona na temat natury działania, która powinna być podstawowym przedmiotem zainteresowania autora.

Niepotrzebne wydaje się także porównanie koncepcji interpretacji radykalnej z teorią strategii intencjonalnej Daniela Dennetta (1987). Pomijając to, że (wbrew m.in. deklaracjom samego Dennetta[1]) Barcz przypisuje Dennettowi instrumentalizm, nie wiadomo, czemu porównanie tych teorii miałoby służyć, skoro koncepcja Dennetta nie zostaje później zastosowana do teorii działania, lecz jedynie (dość autorytatywnie) uznana za „bardziej przekonującą wersję interpretacjonizmu" (Barcz 2020: 53). Do przedstawienia pełnego obrazu brakuje za to omówienia wpływowej, a osadzonej bezpośrednio w dyskursie filozofii działania, krytyki Davidsona autorstwa Judith Thomson (1971) przedstawiającej tzw. „problem czasu wykonania" (*tense problem*) związany z identyfikacją działań ze zdarzeniami fizycznymi i ruchami ciała[2]. Te mankamenty nie przesłaniają jednak tego, że rekonstrukcja poglądów Davidsona dokonana przez Barcza jest skrupulatna i z pewnością dokładniejsza od wielu innych, które można spotkać w literaturze przedmiotu.

Trzeci rozdział skupia się na przedstawieniu koncepcji Melego (1992) i Dretskego (1988) — wpływowych i bardziej współczesnych ujęć teorii przyczynowej, różniących się od Davidsonowskich analiz zarówno naturalistyczną metodologią, jak i celem. Barcz sprawnie relacjonuje poglądy obu autorów (zwracają uwagę użyteczne schematy, które pozwalają zrozumieć mechanizmy opisywane przez Melego i Dretskego). Wskazuje też podstawową lukę obu koncepcji, a mianowicie brak argumentów za ich zgodnością z przewidywaniami nauk empirycznych (neurobiologii w wypadku koncepcji Melego i biologii ewolucyjnej w wypadku teorii Dretskego).

Rozdział ten kończy się jednak dość wątpliwymi uwagami metodologicznymi: teorii Melego Barcz zarzuca, że zajmuje się ona jedynie omówieniem czy rozjaśnieniem potocznych intuicji, a teorii Dretskego — że nie jest w istocie teorią działania, lecz zachowania systemów biologicznych. Według Barcza teoria Dretskego nie przyjmuje mianowicie, że „paradygmatycznymi przypadkami działań są działania intencjonalne" (Barcz 2020: 104). Dlaczego jednak miałoby to dyskwalifikować koncepcję Dretskego? Samo przyjęcie za podstawę pojęcia

---

[1] Por. Dennett 1987: 34-37, 69-81.

[2] Thomson twierdzi, że teoria przyczynowa Davidsona nie radzi sobie z przypadkami działań, w których zaczynający je ruch ciała występuje znacznie wcześniej niż zamierzone konsekwencje (np. sytuacja, w której morderca pociąga za spust, natomiast ofiara ginie kilka dni później — według Thomson Davidsonowska koncepcja wymusza twierdzenie, że zabójstwo, identyfikowane z ruchem palca zabójcy, zaszło wcześniej niż śmierć ofiary). Interesujące rozwinięcie tej dyskusji przedstawia Bennett (1973).

zachowania używanego w etologii przy jednoczesnym wyjaśnieniu intencjonalności działań nie wydaje się wadą *per se*. Dla niektórych podobna cecha mogłaby być zaletą — z pewnością nie sprawia to, że teoria Dretskego nie zalicza się do koncepcji przyczynowych. Jednak nawet jeśli tak jest, to nie jest jasne, dlaczego Barcz decyduje się pisać o wymienionych teoriach tak szczegółowo, zamiast opisać mniej popularne, ale zgodne z jego założeniami metodologicznymi teorie. Podobny problem nierównowagi między omówieniem i rekonstrukcją teoretyczną a kontrargumentacją pojawia się zresztą w tej książce wielokrotnie, do czego przejdę przy omówieniu ostatnich dwóch rozdziałów.

W rozdziale czwartym autor w błyskotliwy sposób kategoryzuje różne teorie analizujące działanie w ramach paradygmatu przyczynowego ze względu na ich założenia ontologiczne i metodologiczne (w szczególności uwzględniając przyjmowany w nich naturalizm lub antynaturalizm metodologiczny). Jest to niezwykle cenny rozdział dla każdego, kto próbuje zrozumieć złożoną materię filozofii działania, na której rozstrzygnięcia mają wpływ rozróżnienia dokonywane w filozofii języka, umysłu, nauki czy ontologii. Analiza Barcza jasno pokazuje oś podziału między różnymi koncepcjami przyczynowymi i unaocznia, jak przyjmowane założenia wpływają na uznanie lub odrzucenie pewnych teorii.

Szczególnie istotne wydaje się tutaj podniesienie kwestii tego, jaki cel przyświeca przyczynowej teorii działania (rozważanej w podrozdziale 4.3, s. 108-117). Czy powinna być ona rekonstrukcją potocznych intuicji (a zatem powiedzieć nam coś o potocznym pojęciu działania czy działania intencjonalnego), czy też raczej zaproponować nowe pojęcia mające zastąpić potoczny dyskurs w naukach przyrodniczych czy kognitywistyce? Czy w ogóle możliwe jest przełożenie naszego dyskursu dotyczącego działań i zdarzeń mentalnych na pojęcia użyteczne naukowo? Odpowiedzi na te pytania Barcz łączy z przyjęciem odpowiednio mocnych pozycji naturalistycznych (od Quine'owskiego programu naturalizacji przez eliminację po umiarkowany naturalizm Bishopa). Rozdział ten zawiera również syntetycznie przedstawioną krytykę Quine'owskiego projektu naturalizacyjnego i płynące z niego wnioski dla metodologii filozofii działania, która według Barcza nie powinna odrzucać potocznych intuicji, lecz dokonywać ich racjonalnej rekonstrukcji (autor wydaje się tutaj bronić klasycznej analizy pojęciowej w odniesieniu do pojęcia działania, choć w ten sposób tego nie charakteryzuje). Wbrew niektórym teoretykom wskazuje, że filozofia powinna wykazać możliwość naturalizacji tych pojęć, a nie ją zakładać (w duchu stwierdzenia: „skomplikowane rzeczy zostawmy do doprecyzowania neurobiologii"). Jest to prawdopodobnie najlepszy rozdział pierwszej części, skutecznie porządkujący wiedzę z zakresu związków filozofii działania z ontologią i filozofią nauki.

2. ARGUMENTY PRZECIWKO PRZYCZYNOWYM TEORIOM DZIAŁAŃ

Po przedstawieniu najpopularniejszych teorii przyczynowych Barcz przechodzi do omówienia historycznie istotnych i wpływowych argumentów przeciwko nim. Rozdział piąty służy rekonstrukcji argumentu ze związku logicznego, który stwierdza, że między intencją działania a jego opisem zachodzi związek konieczności logicznej, gdyż wiedza, że to dana intencja *i* była intencją wykonania pewnego działania *a*, jest dla nas dostępna *a priori*. Ponieważ jednak, zgodnie z Hume'owską analizą przyczynowości, związki przyczynowe są przygodne, zatem *i* nie może być *przyczyną a* (przytaczane jest tutaj klasyczne sformułowanie A. I. Meldena). Barcz przywołuje odparcie tego zarzutu sformułowane przez Davidsona i odwołujące się do spostrzeżenia, że związek konieczności logicznej zachodzi między opisami zdarzeń, a nie między samymi zdarzeniami. Przykładem obrazującym ten kontrargument ma być para zdań:

(*) Zdarzenie A jest przyczyną zdarzenia B

oraz

(ZD) Przyczyna zdarzenia B jest przyczyną zdarzenia B.

Według Davidsona jeśli (*) jest prawdziwe, to prawdziwe jest również (ZD), a (ZD) jest zdaniem analitycznym ze względu na użyte w nim opisy. Zwolennik argumentu ze związku logicznego myli po prostu te różne poziomy opisu, jeden z nich odnosząc do relacji między intencją a działaniem, a drugi — do relacji między zdarzeniami fizycznymi.

Barcz stara się pokazać, że (ZD) nie jest dobrym kontrprzykładem dla argumentu Meldena, ponieważ jest amfiboliczny — jeżeli (ZD) potraktujemy nie jako zdanie stwierdzające tożsamość („Przyczyna zdarzenia B *jest* przyczyną zdarzenia B"), lecz jako zdanie orzekające przyczynowość („Przyczyna zdarzenia B *jest przyczyną* zdarzenia B"), to nie trzeba go uznawać za analityczne. Słusznie i dokładnie wskazuje, że interpretacja (ZD) jako zdania będącego podstawieniem prawa tożsamości jest błędna i że zdanie to należy interpretować w drugi z podanych sposobów. Czy jednak świadczy to przeciwko argumentacji Davidsona? Barcz twierdzi, że nie — jednocześnie nie podaje za tym żadnych racji — pisze jedynie, że „argument Davidsona [...] jest przekonujący. Rozumowania tego nie wspiera jednak przykład (ZD)" (Barcz 2020: 31). Jest to kontrowersyjny fragment, ponieważ Barcz stara się wykazać, że podany kontrprzykład jest błędny, lecz samo rozumowanie jest przekonujące mimo braku jego alternatywnego uzasadnienia.

Warto jednak zauważyć, że replika Barcza wydaje się na pierwszy rzut oka mieszać ze sobą pojęcia analityczności i konieczności. (ZD) według drugiej

interpretacji należałoby odczytać jako dwukrotnie orzekające tę samą własność o pewnym zdarzeniu A („bycie przyczyną zdarzenia B") — aby założyć, że nie jest ono konieczne, należałoby przyjąć, że pierwsze orzeczenie jest użyte referencyjnie, a drugie — atrybutywnie (podobnie do referencyjnego użycia deskrypcji proponowanego przez Keitha Donnellana (1966)). Jest to kwestia sporna, zwłaszcza w wypadku zdań o postaci „[*To*] *F jest F*" (przez wielu uznawanych właśnie za analityczne[3]), czyli takich jak zdanie (ZD) zgodnie z drugim odczytaniem. Jeśli Barcz skłania się ku podobnej interpretacji, powinien ją uzasadnić oraz wskazać, dlaczego mimo wadliwości argumentu rozumowanie Davidsona pozostaje w mocy.

Nie jest to zbyt poważny zarzut w stosunku do książki Barcza, ponieważ rozdział piąty służy głównie rekonstrukcji w dużej mierze historycznego argumentu przeciwko teorii przyczynowej. Dużo ważniejsze dla współczesnej debaty (co słusznie podkreśla autor) są zarzuty dotyczące istnienia dziwnych ciągów przyczynowych (rozdział 6) i tzw. zagubionego sprawcy (rozdział 7).

Rozdział 6 rozpoczyna bardzo istotne dla przebiegu debaty rozróżnienie na wewnętrzne i zewnętrzne dziwne ciągi przyczynowe. Zgodnie z zaproponowaną charakterystyką teorie przyczynowe wyróżnia wspólna teza, że działanie jest definiowalne jako zdarzenie, którego przyczyną jest pewna intencja osoby wykonującej dane działanie. Co jednak wtedy, gdy zdarzenie dochodzi do skutku w sposób niezamierzony? Za Davidsonem Barcz rozróżnia tutaj dwa możliwe przypadki: (1) sytuacje, w których celowe działanie prowadzi do oczekiwanego skutku w niezamierzony sposób (np. działanie mające na celu zabicie Y przez X powoduje wypadek, w którego rezultacie Y ginie bez bezpośredniego udziału X-a; takie wypadki nazywa się zewnętrznymi dziwnymi ciągami przyczynowymi); i (2) sytuacje, w których samo sformułowanie intencji prowadzi do niecelowego zachowania (np. przez wywołanie silnych emocji lub reakcji ciała), którego skutkiem jest zamierzony cel. Zgodnie z teorią przyczynową podobne sytuacje powinny zostać zakwalifikowane jako przypadki działania intencjonalnego, co wyraźnie kłóci się z potoczną intuicją.

To istotne rozróżnienie Barcz jasno stosuje do szeregu przypadków (interesującym spostrzeżeniem autora jest to, że podobne rozważania można zastosować do okazjonalizmu Nicolasa Malebranche'a). Szczególnie ciekawe wydaje się krytyczne zrekonstruowanie odpowiedzi, które na problem dziwnych ciągów przyczynowych podają Mele i John Searle. Mele (1992) stara się odeprzeć problem przez wprowadzenie relacji „bezpośredniego powodowania" przez intencję określonego zdarzenia. Barcz, za Scottem Sehonem (2005), słusznie zauważa, że przyjęcie tego rozwiązania wymaga akceptacji niezwykle

[3] O kontrowersjach dotyczących interpretacji podobnych zdań pisze np. Grudzińska 2007.

wątpliwych założeń empirycznych z zakresu neurobiologii, niepopartych żadnymi dowodami w koncepcji Melego. Autor wskazuje także ciekawy zarzut przeciwko teorii Searle'a (1983), według którego można uniknąć problemu dziwnych ciągów przyczynowych przez założenie, że dane działanie musi przebiegać „ściśle według planu" (Barcz 2020: 141) reprezentowanego przez intencję. Barcz zauważa, że podobne rozumowanie prowadzi do regresu w nieskończoność: problem dziwnych ciągów przyczynowych da się bowiem odtworzyć na dowolnym etapie postępowania zgodnie z planem, teoria Searle'a wymaga więc, by intencja zawierała reprezentację nieskończenie szczegółowego planu. Warto, aby ten klarownie wyłożony zarzut znalazł swoje miejsce w dyskusji dotyczącej koncepcji Searle'a.

Rozdział ten wydaje się jednocześnie najbardziej szczegółowy, ponieważ dokładnie omawia różne kontrpropozycje i ich mankamenty. Bardzo interesujące jest również szerokie wykorzystanie przykładów z zakresu kultury popularnej (np. przypadek bohatera filmu *Memento* Christophera Nolana) czy metody aktorskiej (niezwykle ciekawe odwołanie do metody Stanisławskiego) — pozwala to w angażujący sposób śledzić rekonstrukcję subtelnej pojęciowo debaty. Wszystko to składa się na bardzo syntetyczne i adekwatne streszczenie dyskusji nad dziwnymi ciągami przyczynowymi. Ostatecznie zaś wywód Barcza skłania czytelnika do sceptycznej konkluzji zgodnej z twierdzeniem Davidsona, że istnienie dziwnych ciągów przyczynowych przesądza o tym, że teoria przyczynowa może dostarczyć jedynie koniecznych, lecz już nie wystarczających warunków kwalifikacji danego zdarzenia jako działania.

Z zupełnie innej strony do problemów teorii przyczynowej podchodzą zwolennicy tzw. przyczynowości sprawczej, których zarzuty Barcz rekonstruuje w rozdziale siódmym. Zgodnie z zarzutem pochodzącym od Taylora teoria przyczynowa traktująca działania jako zdarzenia mające swoją przyczynę w określonych zdarzeniach mentalnych nie jest w stanie wyjaśnić wolnego działania samego sprawcy. Wydaje się wręcz dopuszczać, aby „ów ciąg powiązanych przyczynowo zdarzeń wykroczył poza samego sprawcę w ten sposób, że zostanie zapoczątkowany poza sprawcą przez działania kogoś innego […] lub na skutek jakiegoś przypadkowego procesu fizycznego, który będzie indukował nasze intencje" (Barcz 2020: 163). Sprawca nie jest więc aktywny ani nie działa w sposób wolny; można by rzec, że działania *przydarzają mu się* ze względu na formułowane intencje, nie zaś — że jest ich autorem. Zarzut ten cieszy się szczególną popularnością wśród teoretyków przyczynowości sprawczej, którzy uznają go za przesądzający argument przeciwko teoriom przyczynowym. Żeby uznać ów problem za faktyczny kłopot dla teorii przyczynowych, wpierw musimy założyć, że istnieje coś takiego jak „aktywne sprawstwo". Barcz słusznie więc rozpatruje go przez pryzmat formułowanych alternatywnie koncepcji,

które mają posłużyć za lepsze wyjaśnienie pojęcia działania. Zauważa również (w zgodzie ze stosowanym w rozdziale czwartym rozróżnieniem), że podany zarzut może dotyczyć jedynie koncepcji naturalistycznych, takich jak koncepcje Dretskego czy Melego.

Barcz śledzi odpowiedzi sformułowane na gruncie teorii przyczynowości sprawczej, przywołując teorie E. J. Lowe'a (2008), Marii Alvarez i Johna Hymana (1998) oraz Timothy'ego O'Connora (2000), starając się przede wszystkim odpowiedzieć na pytanie stawiane przez wielu przeciwników tych teorii: czy koncepcję przyczynowości sprawczej, przedstawiającą samego sprawcę jako przyczynę zdarzeń, można pogodzić z naukowym obrazem świata? Najistotniejsze staje się tutaj pytanie, czy pojęcie przyczynowości może być stosowane nie tylko do zdarzeń, lecz także do substancji (w tym — wolnego podmiotu) i czy drugi rodzaj przyczynowości może być zredukowany pojęciowo lub ontologicznie do pierwszego. Autorzy postulujący podobne rozwiązanie zazwyczaj odwołują się do analizy przyczynowości jako przejawiania mocy przyczynowych, które to stanowisko zostaje przez Barcza poddane dokładnemu badaniu. Choć brak tu błyskotliwych przykładów z poprzedniego rozdziału, również tutaj klarowność przedstawienia różnych stanowisk przez autora budzi uznanie. Ostatecznie Barcz uznaje — za Charlesem Broadem (1952) — że przyczynowości sprawczej nie można uznać za odmienną ontologicznie od przyczynowości zdarzeniowej, ponieważ jeśli jest ona niedookreślona czasoprzestrzennie, to nie można jej traktować jako realizacji naukowego pojęcia przyczynowości; jeśli zaś można ją w taki sposób dookreślić, to spełnia definicję zdarzenia jako realizacji własności w pewnym miejscu i czasie (por. Kim 1976).

Obszernej analizie Barcza w tym rozdziale można zarzucić tylko to, że zauważając w wielu miejscach, że opisywany problem wiąże się z ogólniejszą debatą dotyczącą spójności między pojęciem wolnej woli a deterministycznym obrazem świata (świadczy o tym choćby omówienie repliki Helen Steward na zarzut Broada), nie podejmuje tego tematu *explicite*. Brak tu przede wszystkim omówienia koncepcji kompatybilistycznych, zgodnie z którymi istnienie wolnej woli nie wyklucza się z deterministycznym obrazem świata, co mogłoby wskazać interesujący kontrargument przeciwko argumentowi z zagubionego sprawcy.

Choć służące rekonstrukcji stanu debaty rozdziały 1-7 nie są pozbawione wad, to autorowi z pewnością nie można zarzucić niedbałości. Dla osoby pragnącej skorzystać z nich jako wprowadzenia do dyskusji nad teorią przyczynową jest to z pewnością rzetelne źródło filozoficznej wiedzy. Już samo to, zwłaszcza ze względu na swoją nowość w polskojęzycznej literaturze, stanowi o bardzo wysokiej wartości *Mechaniki działań*. Osoba czytająca tekst krytycznie prawdopodobnie będzie w stanie szybko odnaleźć źródła czy dotrzeć do nieomówionych dyskusji i samodzielnie wyrobić sobie opinię na temat statusu

teorii przyczynowych. Jedynym problemem może być forma wypowiedzi autora: tekst ma wielopoziomową strukturę i podzielony jest na bardzo dużą liczbę rozdziałów i podrozdziałów różnego poziomu. Może to odstraszać czytelników niezaznajomionych z podobnym stylem (książka oparta jest na rozprawie doktorskiej). Pomijając jednak ten fakt, można z czystym sumieniem polecić tę część *Mechaniki działań* jako pozycję merytorycznie i erudycyjnie przedstawiającą przyczynową teorię działania.

3. DZIAŁANIA OMYŁKOWE I KONEKSJONIZM

Jak jednak zaznaczyłem na wstępie, książka zawiera jeszcze dwa rozdziały, których zadaniem jest przedstawienie oryginalnych zarzutów wobec teorii przyczynowej. Podsumowując dotychczasowe rozstrzygnięcia autora, można zauważyć, że najbardziej przekonującym zarzutem wobec teorii przyczynowej jest problem dziwnych ciągów przyczynowych — okazuje się, że teoria przyczynowa nie dostarcza warunków wystarczających uznania pewnego zdarzenia za działanie. W rozdziale ósmym Barcz podejmuje krytykę tej definicji *jako warunku koniecznego*, przedstawiając zarzut z działań omyłkowych.

Barcz definiuje omyłki, odróżniając je od blisko z nimi spokrewnionych pomyłek, opierając się na charakterystyce Elizabeth Anscombe (1963). Pomyłki to sytuacje, w których skutek działania jest niezgodny z zamierzonym celem, ponieważ opiera się ono na błędnym przekonaniu — na przykład, gdy zamiast masła kupujemy margarynę, ponieważ pomyliliśmy słowa zapisane na liście zakupów. Natomiast omyłki to sytuacje, w których naszemu działaniu nie towarzyszy podobnie błędne przekonanie, lecz gdy nasza intencja powoduje pewne działanie z nią niezgodne. Barcz podaje następujący przykład omyłki:

> Wyobraźmy sobie, że pracuję przy biurku i zauważam, że stojący obok kwiat doniczkowy więdnie. Wstaję więc i ruszam do kuchni po wodę, żeby go podlać. W kuchni odkręcam kran, lecz zamiast podstawić konewkę, odruchowo sięgam po czajnik (zapominam o roślinie) i zaczynam parzyć herbatę. Parę chwil później, gdy stawiam filiżankę herbaty na biurku, uświadamiam sobie, że przecież miałem przynieść wodę (Barcz 2020: 184-185).

W podanej sytuacji możemy powiedzieć, że autor nie tyle *pomylił się* w jakimś sądzie (byłoby tak na przykład, gdyby ktoś poprosił go o podlanie kwiatów, a on rozproszony uznał, że prosi się go o zrobienie herbaty), ile *omyłkowo* zrobił herbatę zamiast podlać kwiaty. O ile pomyłki są więc podatne na pewnego rodzaju racjonalizację (przez wskazanie na błędne przekonania, np. błędną

interpretację prośby lub pytania), o tyle omyłki są działaniami z definicji nieracjonalnymi[4].

Barcz rozważa trzy różne strategie, które mogłyby pozwolić na zasymilowanie omyłek w ramach teorii przyczynowej. Pierwsza polega na próbie teoretycznej redukcji omyłek do pomyłek — a więc wykazania, że w istocie zawsze opierają się one na błędnych przekonaniach czy reprezentacjach podmiotu. Jak słusznie zauważa, w przypadku bardziej złożonych działań podobne rozumowanie wydaje się prowadzić do absurdu, wymaga bowiem szeregu błędnych przekonań i reprezentacji, które wzięte razem, prowadziłyby do twierdzenia, że dany podmiot znajduje się w złożonej halucynacji[5]. Można by powiedzieć, że autor np. miał błędne przekonania co do tego, czy sięga po czajnik czy konewkę, jednak wówczas sytuacja wymagałaby, aby całą złożoną czynność (wyciąganie herbaty, parzenie jej, umycie filiżanki) autor cały czas świadomie uznawał za czynność nalewania wody do konewki. Ta strategia jest więc teoretycznie niepożądana.

Drugim rozwiązaniem może być wyróżnienie dwóch intencji — pierwotnej i niezrealizowanej, inicjującej zachowanie, w toku którego dochodzi do powstania drugiej intencji (nazwijmy ją intencją wtórną), ostatecznie decydującej o działaniu podmiotu. Ze zmiany intencji podmiot ten nie zdaje sobie sprawy; działanie jest jednak intencjonalne (powodowane przez intencję wtórną), choć omyłkowe (o tym charakterze decyduje istnienie niezrealizowanej intencji pierwotnej). Aby móc zaproponować tego typu rozwiązanie, należy wyraźnie scharakteryzować rolę intencji wtórnej, w szczególności odpowiedzieć na pytanie, czy jest ona świadoma, czy nieuświadomiona. Barcz odrzuca pierwszą z możliwości, zauważając, że w takiej sytuacji nie mielibyśmy do czynienia z omyłką — działanie w wyniku intencji wtórnej byłoby po prostu zwykłym nowym działaniem. Rozważając drugą z możliwości, przedstawia rozróżnienie między aktualnie a trwale nieuświadomionymi intencjami. Te pierwsze mogą służyć na przykład wyjaśnieniu działań rutynowych na gruncie teorii przyczynowych. Gdy kieruję samochodem, rzadko formułuję wyraźną świadomą intencję na przykład zmiany biegu lub wciśnięcia pedału gazu — jednakże, jeśli zostanę zapytany o to, *dlaczego* to zrobiłem, jestem

[4] Przez „nieracjonalność" rozumiem tutaj jedynie tzw. nieracjonalność sprawczą (por. Bortolotti 2010), przejawiającą się w niespójności żywionych intencji, przekonań i pragnień z wykonywanymi działaniami. Uznaję, że zarzut Barcza można rozszerzyć na wszelkiego rodzaju działania nieracjonalne sprawczo, ponieważ to właśnie ta cecha omyłek stanowi kłopot dla teorii przyczynowych.

[5] Co interesujące, podobny zarzut względem interpretacjonizmu Dennetta wspominanego wcześniej w książce wysunął Stephen Stich (1981), uznając, że założenie o racjonalności proceduralnej podmiotu w procesie interpretacji przyjmowane przez Dennetta wymaga przypisywania w sytuacjach omyłkowych podobnie nieracjonalnych przekonań.

w stanie udzielić na to pytanie odpowiedzi. Ta cecha pozwala odróżnić podobne zachowania od ruchów wykonywanych na przykład podczas somnambulizmu — jak zauważa Barcz, somnambulicy potrafią wykonywać wiele skomplikowanych czynności, w tym między innymi właśnie prowadzić pojazd. Ponieważ jednak nie określilibyśmy ich zachowań jako działań, podstawowe znacznie ma tu sama *możliwość* dostępu poznawczego.

Wyjaśnienie omyłek przez realizację trwale nieuświadomionych intencji Barcz w interesujący sposób łączy z Freudowską wizją omyłek przedstawioną w tekstach założycielskich psychoanalizy. Rzadko spotyka się w tekstach poświęconych filozofii analitycznej omówienia poglądów Freuda, tym ciekawszy jest więc krótki podrozdział, który zostaje im poświęcony. Wychodząc z założeń wczesnej psychoanalizy, zgodnie z którą istnieje niedostępna nam sfera nieświadomości zawierająca pewne pragnienia i intencje, można by wytłumaczyć działania omyłkowe przez ich realizację. Takie działania byłyby niemożliwe do wytłumaczenia z perspektywy racjonalnej świadomości, byłyby one jednak intencjonalne ze względu na intencje tkwiące w nieświadomości (Barcz barwnie ilustruje to przykładem z dzienników Freuda). Autor odrzuca jednak Freudowskie wyjaśnienie, przywołując zarzut Poppera wobec psychoanalizy jako niefalsyfikowalnej empirycznie. Jest to dość zaskakujące, gdyż mimo swojej sławy Popperowska krytyka psychoanalizy była wielokrotnie atakowana zarówno ze strony empirycznej (por. Grant, Harari 2005), jak i filozoficznej (por. Grünbaum 1979) i raczej nie jest uznawana za merytoryczną.

Freudowskie wyjaśnienie omyłek jako zgodne z teorią przyczynową można by, moim zdaniem, odeprzeć znacznie lepiej, opierając się na wspomnianych wcześniej przykładach zachowań nieintencjonalnych takich jak somnambulizm czy stany nieświadomości wywołane chorobą lub upojeniem. Posługując się psychoanalitycznym aparatem w wersji przedstawionej przez Barcza, nie możemy bowiem wykluczyć tych przypadków z klasy działań intencjonalnych, ponieważ w podobnych wypadkach można wytworzyć odpowiedniego rodzaju wyjaśnienie w kategoriach nieświadomej intencji (zważywszy choćby na rolę, którą według Freuda odgrywały sny jako przestrzeń oddziaływania nieświadomości). Nie jest to więc, jak się zdaje, teoria użyteczna dla zwolennika teorii przyczynowej, ponieważ jej przyjęcie wymagałoby porzucenia intuicyjnego podziału na działania i niedziałania[6]. Szkoda jednak, że Barcz decyduje się na skwitowanie długo omawianej przez siebie i twórczo rekonstruowanej koncepcji dwoma akapitami podnoszącymi obiekcję wątpliwej klasy, która nie zostaje odpowiednio rozbudowana.

---

[6] Podobny zarzut pojawia się zresztą później, kiedy Barcz omawia koncepcję, zgodnie z którą omyłki to działania powodowane przez doznania (Barcz 2020: 212).

Barcz uzasadnia jednak odrzucenie wyjaśnienia w kategoriach aktualnie nieuświadomionych intencji w sposób jeszcze bardziej oględny, pisząc jedynie, że „pojęcie intencji aktualnie nieuświadomionej nie może […] służyć do wykazania intencjonalnego charakteru omyłek z zachowaniem ich *omyłkowego* charakteru" (Barcz 2020: 205). Ale właściwie dlaczego? Wróćmy do przykładu omawianego przez Barcza, czyli sytuacji, w której decyduje się on podlać kwiaty, lecz gdy idzie do kuchni, omyłkowo zaczyna parzyć herbatę zamiast napełnić konewkę. Czy nie dałoby się wyjaśnić tego działania właśnie analogicznie do zachowania rutynowego, gdzie w odpowiednim środowisku (kuchnia) dochodzi do wytworzenia aktualnie nieuświadomionej intencji (zaparzenia herbaty), zgodnie z którą należy wykonać szereg czynności (wstawienie czajnika, zalanie torebki z herbatą itp.)? Możemy z łatwością wyobrazić sobie, że przechodząca osoba zadaje autorowi pytanie: „co robisz?", na które ten odpowiada: „parzę herbatę" (aktualizując intencję wtórną), i dopiero później zdaje sobie sprawę, że miał podlać kwiaty (a więc uświadamia sobie intencję pierwotną, która czyni to wydarzenie omyłką). Czy w takim wypadku nie jest tak, że spełniamy zarówno wymagania teorii przyczynowej, jak i nadajemy temu działaniu charakter omyłkowy?

Podobnie rzecz ma się w wypadku trzeciej strategii, polegającej na zinterpretowaniu omyłek jako powodowanych przez doznania. Strategia ta zostaje podsumowana jedynie w półtorastronicowym podrozdziale (8.3.4.6). Autor decyduje się odeprzeć tę koncepcję przez wskazanie, że również takie zachowania jak somnambulizm są wrażliwe na doznania, wobec czego, wbrew naszej intuicji, powinny one zostać uznane za działania. Czemu jednak nie uczynić tutaj podobnego zastrzeżenia co w przypadku intencji, a mianowicie — że przyczynami działań omyłkowych mogą być jedynie doznania, które można sobie uświadomić? Podobne wyjaśnienia wydają się zresztą zgodne z naszą potoczną praktyką. Wróćmy do przykładu przytaczanego przez Barcza. Załóżmy, że osoba, która usiądzie przy stole z kubkiem zaparzonej herbaty i zda sobie sprawę, że faktycznie miała zamiar podlać kwiaty, zada sobie pytanie: „dlaczego właściwie to zrobiłem?". Naturalne byłoby powiedzenie: „musiałem to zrobić dlatego, że zobaczyłem czajnik i odruchowo postanowiłem zrobić herbatę" (wskazując na doznanie, jakim jest percepcja czajnika) albo „żeby podlać kwiaty, musiałem pójść do kuchni, a przecież w niej zawsze robię herbatę" (doznanie przebywania w kuchni). Osobiście uznaję podobne wyjaśnienie za przekonujące i nie dostrzegam w książce Barcza dobrego kontrargumentu przeciwko takiemu ujęciu przypadków omyłek.

W podsumowaniu tego rozdziału Barcz pisze, że omyłki podlegają pewnego rodzaju wyjaśnieniom przyczynowym, jednak nie w rozumieniu akceptowalnym przez teorię przyczynową. Ma tutaj na myśli czynniki takie jak zmęczenie,

roztargnienie czy stres — tych stanów psychicznych używamy najczęściej, gdy musimy wyjaśnić *przyczyny* omyłki. Należy jednak zauważyć, że wymienione przez Barcza stany nie są w pełnym tego słowa znaczeniu *przyczynami* określonych zdarzeń — są raczej okolicznościami sprzyjającymi, nie zaś *zdarzeniami*, które moglibyśmy uznać za pełnoprawną przyczynę występowania konkretnej omyłki. Stres, w wyniku którego wystąpiła omyłka, nie jest zdarzeniem — może nim być dopiero błędne spostrzeżenie lub pewne doznanie. Niestabilna konstrukcja mostu nie jest zdarzeniem, choć z pewnością sprzyja jego zawaleniu — może być ono jednak *spowodowane* dopiero przez konkretne przeciążenie lub burzę. Podobne spostrzeżenie czyniłoby również bardziej prawdopodobną analogię między omyłkami a działaniami rutynowymi. Osoby będące pod wpływem tych czynników częściej polegać będą na automatyzmach czy często powtarzanych przez siebie czynnościach, których procedurę mają wpisaną w rutynę.

Choć uważam argumenty przedstawione przez Barcza za niekonkluzywne, to z pewnością zwrócił uwagę na istotny problem. Wiele teorii przyczynowych *explicite* (jak teoria Davidsona) czy *implicite* (jak teorie Melego czy Searle'a) przyjmuje, że wszelkie działania są nie tylko intencjonalne, lecz także racjonalne. Wiele jednak z naszych zachowań, które bez większych oporów uznalibyśmy za działania, nie poddaje się opisowi w kategoriach racjonalnych. Należy więc uważnie przyjrzeć się koncepcji racjonalności zakładanej przez te teorie[7] — jest to z pewnością istotny wniosek.

Ostatni rozdział został poświęcony równie ciekawemu zagadnieniu, a mianowicie zgodności między różnymi teoriami działania a paradygmatami badań nad sztuczną inteligencją. Ta dziedzina jest jednocześnie ważnym polem zastosowań różnych koncepcji działania i głębokim źródłem empirycznej wiedzy dotyczącej zachowania systemów poznawczych. Cieszy więc, że Barcz porusza tę kwestię; szkoda jednak, że spośród wszystkich rozdziałów to właśnie ten wydaje się najbardziej niedopracowany metodologicznie. Fakt, że tekst odwołuje się tylko do ośmiu nowych źródeł (wliczając w to obszerny cytat z *Lewiatana* Thomasa Hobbesa), z czego tylko dwa są tekstami napisanymi po 2000 roku, każe bowiem podejść sceptycznie do wartości merytorycznej rozdziału dotyczącego dynamicznie rozwijającej się dziedziny.

Niemal połowę rozdziału zajmuje omówienie głośnej krytyki Huberta Dreyfusa wymierzonej w mocny obliczeniowy program sztucznej inteligencji

[7] Przykładami takich działań mogą być zachowania osób z urojeniami. Czy osobę, która przestaje jeść, ponieważ uznaje się za martwą (jak w wypadku zespołu Cotarda), można uznać za racjonalną sprawczo? Podobne przykłady wydają się kłopotliwe zwłaszcza dla teorii Davidsona, łączącego wyjaśnienie zachowania z przypisywaniem przekonań i postulującego założenie o racjonalności proceduralnej zbioru przekonań (por. Tarnowski 2019).

(uznający tożsamość myślenia z systemem przetwarzania symbolicznego). Warto jednak podejść do niej sceptycznie. Po pierwsze dotyczy ona dość jednoznacznie trendu w badaniach nad sztuczną inteligencją (czyli mocnego komputacjonizmu spod znaku programu Herberta Simona i Alana Newella) rozwijanego w latach 50. i 60. XX wieku — a więc ponad pół wieku temu. Sama recepcja odpowiedzi Dreyfusa, choć początkowo nieprzychylna, spowodowała również znaczące zmiany w samym paradygmacie obliczeniowym. Wbrew twierdzeniom Barcza, w obrębie komputacjonizmu doszło do przeobrażeń mających na celu wzięcie pod uwagę niektórych zarzutów Dreyfusa — w obecnej postaci pozostaje on wciąż niezwykle istotnym czy nawet dominującym poglądem w sztucznej inteligencji[8].

Barcz za Dreyfusem uznaje, że za kłopotami teorii klasycznej stoi pewne założenie, które określa mianem *tezy o eksplicytności umysłu* (Barcz 2020: 226). Zgodnie z nią wszystkie działania wykonywane przez umysł muszą mieć sformułowaną *explicite* reprezentację (instrukcję) — teza ta wynika wprost z założenia o algorytmiczności działania umysłu. Jak jednak ma się ta teza do przyczynowej teorii działania? Barcz stwierdza, że zachodzi tutaj analogia między *explicite* reprezentowanymi instrukcjami a intencjami, które według niektórych teoretyków, np. Searle'a czy Bratmana (1984), reprezentują również plany działania. Jak twierdzi, istnienie intencji nieuświadomionych nie stoi w sprzeczności z tezą o eksplicytności, ponieważ teza ta nie stwierdza, że te instrukcje muszą być aktualnie uświadomione w trakcie wykonywania obliczenia. Jest to z pewnością ciekawa analogia, zastanawia jednak, czy sięga tak daleko, jak pragnąłby autor. Jeden z głównych zarzutów Dreyfusa dotyczy niebrania pod uwagę procesów nieświadomych i kontekstu, zwłaszcza w wypadku działań takich jak poruszanie się, rozpoznawanie wzorców itp. (zarzut ten przywołuje Barcz: 234-235). Można wątpić, czy zarzut ten rzeczywiście godzi w sztuczną inteligencję rozwijaną przez Simona i Newella (por. Miłkowski 2013: 181-183). Warto jednak zastanowić się, czy w ogóle taką właśnie eksplicytną wizję procesów umysłowych zakłada teoria przyczynowa. Czy działania takie jak podniesienie przeze mnie ręki czy przejście z jednego końca korytarza na drugi wymagają według Searle'a[9] czy Bratmana reprezentacji w postaci

[8] Szerokie omówienie obliczeniowej teorii umysłu i jej ograniczeń ze współczesnej perspektywy zawiera np. Miłkowski 2013.

[9] Warto też zauważyć, że sam Searle jest autorem dyskutowanych do dziś argumentów przeciwko istnieniu silnej sztucznej inteligencji w jej wersji postulowanej przez komputacjonizm (Searle 1980). Rozważenie, czy stanowisko Searle'a w filozofii umysłu i filozofii działania jest w tym aspekcie spójne, wykracza daleko poza ramy mojej recenzji, jednak ciężar dowodu tego, że pogląd Searle'a na naturę intencji jest bezpośrednio powiązany z obliczeniową wizją umysłu, wydaje się leżeć po stronie Barcza.

drobiazgowego planu kolejnych ruchów mięśni? Wydaje się, że rozumienie „planu" przyjmowane przez teoretyków rozwijających koncepcję przyczynową nie jest równie drobnoziarniste, wobec czego analogia między reprezentacjami planów działania a eksplicytną reprezentacją procesów w komputacjonizmie nie sięga tam, gdzie faktycznie stanowi to problem według Dreyfusa. Trudno więc stwierdzić, na ile krytyka klasycznego komputacjonizmu uderza również w teorię przyczynową.

Następnie Barcz zdaje się wyraźnie zmieniać strategię[10] ataku: wskazuje bowiem na ontologiczną podstawę teorii przyczynowej, którą jest przyjęcie identyczności zdarzeń mentalnych ze zdarzeniami fizycznymi (czy to ich typami, czy egzemplarzami). Wobec tego założenia rekonstruuje dwa argumenty — pierwszy, argument Jennifer Hornsby (1997), dotyczy samej teorii identyczności typów, drugi to przywoływany za Sehonem argument Williama Ramseya, Stephena Sticha i Josepha Garona (1991), który postuluje niezgodność realistycznie rozumianego istnienia nastawień sądzeniowych z paradygmatem koneksjonistycznym w sztucznej inteligencji. Choć trzeba zauważyć, że oba argumenty znajdowały swoje odpowiedzi, których autor nie stara się odeprzeć ani nawet o nich nie wspomina (por. np. Papineau 2007, Clark 1995, Skokowski 2009), to warto zastanowić się nad samą argumentacją proponowaną przez Barcza. Twierdzi, że teoria przyczynowa rzeczywiście wymaga logicznie przyjęcia tezy o identyczności typów stanów mentalnych z typami zdarzeń fizycznych. Podobne rozumowanie wymagałoby jednak przyjęcia zasady domknięcia przyczynowego, zgodnie z którą wszystkie zdarzenia fizyczne mają tylko i wyłącznie fizyczne przyczyny. Barcz wydaje się wyraźnie iść tropem Jaegwona Kima (1989), wskazując, że przyjęcie tej zasady oraz odrzucenie epifenomenalizmu wymagają przyjęcia jakiejś formy teorii identyczności typów, jeśli chcemy utrzymywać, że zdarzenia mentalne w ogóle mogą wchodzić w relacje przyczynowe ze zdarzeniami fizycznymi. Jest to jednak problematyczne rozumowanie, jeśli ma służyć podważeniu teorii przyczynowej jako całości. Uderza ono co najwyżej w fizykalistyczne przyczynowe teorie działania — to one bowiem będą akceptować zasadę domknięcia przyczynowego. Nie jest więc kłopotliwe dla stanowisk antynaturalistycznych (jak teoria Davidsona) ani takich, które

[10] Co należy w tym punkcie odnotować, poglądy Dreyfusa istotne są dla innego paradygmatu obecnego w badaniach nad sztuczną inteligencją i kognitywistyce, czyli enaktywizmu (Varela, Thompson, Rosch 2017), który nie jest bezpośrednio związany z koneksjonizmem, ale Barcz w ogóle go nie omawia. Sam Dreyfus w odniesieniu do odrodzenia koneksjonizmu w latach 90. XX wieku w drugim wydaniu swojej książki pisze: „Jest wielce prawdopodobne, że odrzucone i przeżywające swoje odrodzenie podejście koneksjonistyczne otrzymuje jedynie swoją zasłużoną szansę na porażkę" (Dreyfus 1992: xxxviii). Warto więc zastanowić się, czy wiele spośród jego krytyki, jeśli już decydujemy się ją przyjąć, nie znajduje swojego zastosowania również wobec tego podejścia do sztucznej inteligencji.

odżegnując się od fizykalizmu, pozostają teoriami naturalistycznymi (jak koncepcja Searle'a czy inne postulujące dualizm własności w odniesieniu do umysłu). O ile więc przytaczane argumenty są słuszne, uderza to tylko w wybrane teorie przyczynowe. Nie ma w tym jednak nic dziwnego, że argumenty przeciwko fizykalistycznej identyczności typów uderzają w teorie, które ją *explicite* zakładają. Barcz niestety nie tłumaczy, dlaczego ta krytyka ma, jego zdaniem, szerszy zasięg.

ZAKOŃCZENIE

Na koniec przejdźmy do kwestii metodologicznych. Niezależnie od tego, jak ocenimy zarzuty formułowane przez Barcza, treść *Mechaniki działań* skłania do pytania: czy wobec tak licznych problemów paradygmat przyczynowy powinien być dalej rozwijany? Czy powinniśmy oczekiwać choćby rozwiązania problemu dziwnych ciągów przyczynowych w stopniu pozwalającym dalej utrzymywać teorię przyczynową, czy też rozpocząć poszukiwania alternatywnego modelu ludzkiego działania? W ostatnim akapicie książki, po wyliczeniu ustaleń przywołanych w kolejnych rozdziałach, Barcz konkluduje:

> teoria kauzalna wydaje się dostarczać zasadniczo trafnego obrazu działania, jednak wrażenie to jest mylne. […] [W świetle ustaleń dokonanych w pracy] teoria ta — przynajmniej w znanych z literatury sformułowaniach — nie stanowi satysfakcjonującej analizy ludzkiego działania (Barcz 2020: 254).

Ta konkluzja wydaje się ogólnie trafna — biorąc choćby pod uwagę niezadowalające odpowiedzi na problem dziwnych ciągów przyczynowych czy wątpliwe założenie ludzkiej racjonalności sprawczej obecne w tych teoriach. Jaką propozycję alternatywną możemy jednak znaleźć? Najpopularniejsza z konkurencyjnych koncepcji, czyli teoria przyczynowości sprawczej, również spotyka się na stronach tej książki z zasłużoną krytyką (rozdział siódmy). Obraz problemów nie składa się w żadną spójną całość — książka służy raczej filozoficznej ocenie wielu różnych podejść teoretycznych do teorii przyczynowej oraz zarzutów pod jej adresem. Można jednak spróbować domyślić się ogólnej intuicji, która może stać za silnym odwołaniem do eliminatywizmu metodologicznego w filozofii umysłu w rozdziale dziewiątym czy sugestią wyjaśnień przyczynowych w kategoriach stanów psychicznych w rozdziale ósmym. Pomocnym tropem może być także sam tytuł odnoszący się do ścisłego powiązania filozoficznego pojęcia działania z naukami przyrodniczymi. Sugerowaną drogą może być więc właśnie próba sformułowania definicji działania oraz działania intencjonalnego w sposób, który unikałby pojęć psychologii potocz-

nej. Prowadziłoby to do podania empirycznie sprawdzalnych praw psychologicznych rządzących działaniem — a więc tytułowej „mechaniki działań".

Pewnym kłopotem dla takiego projektu pozostaje jednak metodologiczne pytanie dotyczące treści pojęcia działania. Szczególnie interesujący w tym kontekście wydaje się fakt, że proponowany na końcu książki eliminatywizm wobec nastawień sądzeniowych pociąga za sobą również zaprzeczenie istnieniu intencji, czyli bytu, bez którego trudno wyobrazić sobie adekwatną rekonstrukcję pojęcia działania intencjonalnego. Barcz proponuje jako możliwe wyjście dla nieeliminatywisty przyjęcie instrumentalizmu w stylu Dennetta (Barcz 2020: 245-246), twierdząc jednak, że instrumentalizm wyklucza się z tezą o przyczynowym oddziaływaniu stanów mentalnych, a więc i z przyczynową teorią działania. Czym jednak miałyby być przekonania, intencje czy pragnienia, jeśli zostałyby wykluczone z wyjaśniania i przewidywania ludzkiego działania — co wydaje się podstawowym celem psychologii potocznej? Jeśli nie miałoby to być wyjaśnienie w kategoriach przyczynowych (obecne przecież w naszej praktyce językowej), to jak inaczej wytłumaczyć ich rolę teoretyczną? Sam Dennett zwraca uwagę przede wszystkim na ten aspekt w swojej teorii strategii intencjonalnej — nastawienia sądzeniowe służą nam do tego, aby wyjaśniać, przewidywać i tłumaczyć ludzkie zachowanie (a czynimy to właśnie w sposób przyczynowy: „Zrobił tak, *bo* myślał tak-i-tak", „Takie-i-takie jego zachowanie *było spowodowane* takim-a-takim pragnieniem"). Pojęcie działania w głęboki sposób połączone jest z innymi pojęciami psychologii potocznej i możliwe nawet, że samo do niego należy. Być może więc przewrotną konkluzją z lektury książki Barcza powinien być nie tyle sceptycyzm dotyczący przyczynowej teorii działania, ile sceptycyzm co do samej próby spójnej filozoficznej konstrukcji pojęcia działania.

Pracę *Mechanika działań* Michała Barcza warto zatem rozpatrywać na kilku różnych płaszczyznach. Pod względem merytorycznym jest to książka bogata, stanowiąca bardzo dobre wprowadzenie do problematyki współczesnej analitycznej filozofii działania. Z pewnością jej lektura pozwoli wielu osobom na zrozumienie tej wciąż zbyt mało popularnej w Polsce dziedziny i może być ona bardzo dobrym polskojęzycznym źródłem akademickim. Choć w mojej opinii część krytyczna *Mechaniki działań* nie dostarcza wystarczających argumentów przeciwko przyczynowej teorii działania, z pewnością odnosi się do ważnych, a wciąż zbyt mało badanych w literaturze przedmiotu problemów. Dlatego każdy, kto chce poznać tę dziedzinę lub podjąć w niej samodzielne badania, powinien książkę Barcza przeczytać.

BIBLIOGRAFIA

Alvarez M., Hyman J. (1998), *Agents and Their Actions*, „Philosophy" 73 (2): 219-245.
Anscombe G. E. M. (1963), *Intention*, Oxford: Blackwell.
Antony L. (1989), *Anomalous Monism and the Problem of Explanatory Force*, „The Philosophical Review" 98(2), 153-187. https://doi.org/10.2307/2185281
Barcz M. (2020), *Mechanika działań. Filozoficzny spór wokół przyczynowej teorii działania*, Warszawa: Wydawnictwo Naukowe PWN.
Bennett J. (1973), *Shooting, Killing and Dying*, „Canadian Journal of Philosophy" 2(3), 315-323. https://doi.org/10.1080/00455091.1973.10716046
Bortolotti L. (2010), *Delusions and Other Irrational Beliefs*, Oxford: Oxford University Press. https://doi.org/10.1093/med/9780199206162.001.1
Bratman, M. (1984), *Two Faces of Intention*. „The Philosophical Review" 93(3), 375-405. https://doi.org/10.2307/2184542
Broad Ch. D. (1952), *Ethics and the History of Philosophy,* London: Routledge.
Clark A. (1995), *Connectionist Minds* [w:] *Connectionism: Debates on Psychological Explanation*, C. MacDonald, G. MacDonald (eds.), Oxford: Blackwell, 339-356.
Davidson D. (1992), *Eseje o prawdzie, języku i umyśle*, tłum. B. Stanosz, Warszawa: Wydawnictwo Naukowe PWN.
Dennett D. (1987), *The Intentional Stance*, Cambridge, MA: MIT Press.
Donnellan K. S. (1966), *Reference and Definite Descriptions*, „The Philosophical Review" 75(3), 281-304. https://doi.org/10.2307/2183143
Dretske F. (1988), *Explaining Behavior. Reasons in a World of Causes*, Cambridge, MA: MIT Press.
Dreyfus H. L. (1992), *What Computers Still Can't Do: A Critique of Artificial Reason*, Cambridge, MA: MIT Press.
Grant D. C., Harari E. (2005), *Psychoanalysis, Science and the Seductive Theory of Karl Popper*, „Australian & New Zealand Journal of Psychiatry" 39(6): 446-452. https://doi.org/10.1080/j.1440-1614.2005.01602.x
Grudzińska J. (2007), *Semantyka nazw jednostkowych*, Warszawa: Wydawnictwo Naukowe Semper.
Grünbaum A. (1979), *Is Freudian Psychoanalytic Theory Pseudo-scientific by Karl Popper's Criterion of Demarcation?*, „American Philosophical Quarterly" 16(2), 131-141.
Hornsby J. (1997), *Simple Mindedness: In Defense of Naive Naturalism in the Philosophy of Mind,* Cambridge, MA: Harvard University Press.
Kim J. (1976), *Events as Property Exemplifications* [w:] *Action Theory*, Dordrecht: Springer. https://doi.org/10.1007/978-94-010-9074-2_9
Kim J. (1989), *The Myth of Non-reductive Materialism*, „Proceedings and Addresses of the American Philosophical Association" 63(3): 31-47. https://doi.org/10.2307/3130081
Lowe E. J. (2008), *Personal Agency. The Metaphysics of Mind and Action*, Oxford: Oxford University Press.
Mele A. R. (1992), *Springs of Action,* Oxford: Oxford University Press.
Miłkowski M. (2013), *Explaining the Computational Mind*, Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9339.001.0001
O'Connor T. (2000), *Persons and Causes: The Metaphysics of Free Will*, Oxford: Oxford University Press.

Papineau D. (2007), *Naturalism* [w:] *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), E. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/naturalism/>.

Ramsey W., Stich S., Garon J. (1991), *Connectionism, Eliminativism and the Future of Folk Psychology* [w:] *The Future of Folk Psychology, Intentionality and Cognitive Science*, J. Greenwood (ed.), Cambridge University Press, Cambridge.

Searle J. R. (1980), *Mind, Brains and Programs: A Debate on Artificial Intelligence*, „The Behavioral and Brain Science" 3, 128-135. https://doi.org/10.1017/S0140525X00005756

Searle J. R. (1983), *Intentionality: An Essay in the Philosophy of Mind,* Cambridge: Cambridge University Press.

Sehon S. (2005), *Teleological Realism*, London: MIT Press.

Skokowski P. (2009), *Networks with Attitudes*, „AI & Society" 10, 461-470. https://doi.org/10.1007/s00146-007-0175-5

Stich S. P. (1981), *Dennett on Intentional Systems*, „Philosophical Topics" 12(1), 39-62. https://doi.org/10.5840/philtopics198112142

Tarnowski M. (2019), *Czy posiadanie sprzecznych przekonań jest możliwe? Omówienie i krytyka argumentów za psychologiczną zasadą niesprzeczności*, „Studia Semiotyczne" 33(2), 323-353.

Thomson J. J. (1971), *The Time of a Killing*, „The Journal of Philosophy" 68(5): 115-132. https://doi.org/10.2307/2025335

Varela F. J., Thompson E., Rosch E. (2017), *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262529365.001.0001