

Katarzyna Paprzycka

O argumentacji Lowe'a na rzecz koncepcji przyczynowości intencjonalnej i fizycznej

Filozofia Nauki 21/1, 91-112

2013

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Katarzyna Paprzycka

O argumentacji Lowe'a na rzecz koncepcji przyczynowości intencjonalnej i fizycznej¹

Lowe (2008) rozwija swą filozoficznie interesującą koncepcję niekartezjańskiego dualizmu substancji (por. też Lowe 1996), zgodnie z którą istnieją co prawda dwa rodzaje substancji (fizyczna i mentalna), jest też tak jak sądził Kartezjusz, a mianowicie, że tylko substancja mentalna (podmiot, *ja*) posiada własności mentalne, ale nie jest prawdą, że podmiot nie posiada własności fizycznych. Podmioty mają ciała (choć nie są z nimi tożsame) i w związku też z tym posiadają własności fizyczne.

W ramach tej ogólnej koncepcji Lowe broni kontrowersyjnego poglądu na naturę procesów przyczynowych zachodzących w wypadkach dobrowolnego działania podmiotu (por. szczególnie Lowe 2006 oraz 2008, rozdział 5). W niniejszym tekście zajmiemy się wyłącznie tym fragmentem koncepcji Lowe'a. Autor ten twierdzi, że w wypadkach działania podmiotu mamy do czynienia z dwoma równoległymi, uzupełniającymi się procesami przyczynowymi. Jeden z tych procesów to proces przyczynowania intencjonalnego, a drugi to proces przyczynowania fizycznego. Przyczyną mentalną jest decyzja sprawcy (a tej z kolei sprawca²), a przyczyną fizyczną — złożone zdarzenie neuronalne. Co więcej, Lowe twierdzi, że przyczyny mentalne oddziałują inaczej niż przyczyny fizyczne. Przyczyny mentalne determinują to, że ruch ciała w ogóle ma miejsce (to, że ręka się unosi), natomiast przyczyny neuronalne określają już tylko detaliczne parametry tego ruchu (to, że ręka unosi się z dokładną taką prędkością, po dokładnej takiej trajektorii, w dokładnym takim czasie).

Koncepcja Lowe'a jest nowatorska i z pewnością znajdzie licznych zwolenników, a także przeciwników. Pomysł, że przyczyny neuronalne determinują tylko

¹ Chciałabym podziękować anonimowemu recenzentowi za szereg uwag krytycznych.

² Lowe jest zwolennikiem teorii kauzalności sprawczej, ale nie będzie ona dla nas istotna.

drobiazgowe parametry ruchu, natomiast nie determinują tego, że do ruchu dochodzi, może wydać się na pierwszy rzut oka niezrozumiały. (To trochę tak, jak gdyby powiedzieć, że wzrost temperatury zasadniczo powoduje to, że w ogóle dochodzi do wzrostu ciśnienia gazu, natomiast zupełnie *niezależnie* wzrost energii kinetycznej cząsteczek determinuje już tylko w sposób drobiazgowy zachowanie cząsteczek gazu, nie determinując jednak samego wzrostu ciśnienia.) Istnieją dwa podstawowe źródła tej myśli Lowe'a. Po pierwsze, twierdzi, że odwołanie do zdarzeń neuronalnych nie jest w stanie wyjaśnić tego, że do ruchu ręki w ogóle dochodzi, ponieważ zdarzenia neuronalne są tzw. przyczynami koincydentalnymi. Po drugie, opiera się on na argumente z implikacji kontrfaktycznych, zgodnie z którym niewystąpienie przyczyny mentalnej prowadziło do niewystąpienia ruchu ręki, natomiast niewystąpienie przyczyny neuronalnej prowadziło do wystąpienia nieznacznie innego ruchu ręki. Podstawowym celem tego tekstu jest przede wszystkim pokazanie, że argumentacja Lowe'a jest nieprzekonująca. Argument z implikacji kontrfaktycznych jest błędny (§3-§5), natomiast odwołanie się do przyczyn koincydentalnych jest dyskusyjne (§1).

Rozpoczniemy od naszkicowania stanowiska Lowe'a w kwestii różnic między wyjaśnieniami intencjonalnymi i neuronalnymi, a także domniemanymi — odpowiadającymi tym wyjaśnieniom — procesami przyczynowymi (§1). Wskażemy też, jaką rolę odgrywa argument z implikacji kontrfaktycznych, który zrekonstruujemy w §2. W §3 przedstawimy dwa kontrprzykłady dla stosowanej przez Lowe'a argumentacji, by następnie (§4-§5) przyjrzeć się źródłom problemów.

1. RÓŻNICA MIĘDZY WYJAŚNIENIAMI INTENCJONALNYMI I NEURONALNYMI

Załóżmy, że Iksiński uniosł rękę, gdyż zdecydował się zagłosować na Igrekowską. Lowe zwraca uwagę, że powołanie się na decyzję Iksińskiego nie wyjaśnia, dlaczego jego ręka uniosła się w dokładnie taki sposób (w dokładnie tym czasie, po dokładnie takiej trajektorii ruchu, z dokładnie taką prędkością etc.). Wyjaśnienie to wystarczy jedynie do zgrubnego określenia ruchu ręki — powoływanie się na decyzję nie byłoby adekwatne, gdyby ręka Iksińskiego uniosła się rok po spotkaniu wyborczym, gdyby trajektoria ruchu ręki była zbyt „dziwaczna” (wyobraźnię pobudzić tu mogą skecze Jasia Fasoli) czy też gdyby jej unoszenie odbywało się np. w ślimaczym tempie etc. Lowe twierdzi, że detaliczne parametry uniesienia ręki mogą zostać wyjaśnione przez odwołanie się do pewnego zdarzenia neuronalnego.³

³ Myśl ta (że wyjaśnienia intencjonalne nie wyjaśniają szczegółowych parametrów fizycznych ruchu ciała) pojawia się też np. u von Wrighta (1971). Odrzucają ją natomiast inni zwolennicy teleologicznej koncepcji wyjaśniania działań odwołujący się do intencji *de re* (por. np. Wilson 1989).

Te twierdzenia Lowe'a niektórym mogą wydać się dyskusyjne⁴, ale wzbudzone tymi tezami wątpliwości są raczej niszowe. Rzeczywiście kontrowersyjne twierdzenia Lowe'a dotyczą przede wszystkim tezy o pełnej asymetrii tych wyjaśnień oraz przeniesienia tez o wyjaśnieniach na tezy o przyczynowości. Przyjrzyjmy się obu tym kwestiom.

Jak widzieliśmy, Lowe przyjmuje trzy mniej kontrowersyjne twierdzenia dotyczące wyjaśnień, a także towarzyszące im twierdzenia ontologiczne:

- (M+) Wyjaśnienie intencjonalne (tu: odwołujące się do decyzji podmiotu, aby κ -ować) wyjaśnia, dlaczego w ogóle doszło do κ -owania.
- (OM+) Przyczyny mentalne determinują to, że w ogóle doszło do κ -owania.
- (M-) Wyjaśnienie intencjonalne nie wyjaśnia fizycznych parametrów κ -owania (poza wyznaczeniem pewnych zgrubnych ich ograniczeń).
- (OM-) Przyczyny mentalne nie determinują fizycznych parametrów κ -owania (poza wyznaczeniem pewnych zgrubnych ograniczeń).
- (N+) Wyjaśnienie neuronalne (tu: odwołujące się do przyczyny neuronalnej κ -owania) wyjaśnia wszystkie fizyczne parametry κ -owania.
- (ON+) Przyczyny neuronalne determinują wszystkie fizyczne parametry κ -owania.

Do tych twierdzeń Lowe dodaje jednak jeszcze jedną parę tez, uzyskując w ten sposób pełną asymetrię wyjaśnień intencjonalnych i neuronalnych, a także przyczyn intencjonalnych i neuronalnych, a mianowicie:

- (N-) Wyjaśnienie neuronalne nie wyjaśnia, dlaczego w ogóle doszło do κ -owania.
- (ON-) Przyczyny neuronalne nie determinują tego, że doszło do κ -owania.

Z ontologicznego punktu widzenia wydaje się to koncepcja dyskusyjna, która musiałaby zostać poddana gruntownej analizie, czego nie będę się tutaj podejmować. Chcę zwrócić uwagę na dwie kwestie. Po pierwsze, nie są jasne podstawy dla przyjęcia tez ontologicznych.⁵ *Prima facie* jednak obserwacja, że wyjaśnienie odwołujące się do decyzji podniesienia ręki nie wyjaśnia, dlaczego uniesienie ręki miało dokładnie takie, a nie inne parametry fizyczne, jest obserwacją dotyczącą *wyjaśnienia* —

⁴ Eliminatywiści np. będą podważać tezę, że wyjaśnienia intencjonalne rzeczywiście cokolwiek wyjaśniają, bo powołują się one na nieistniejące (w ich przekonaniu) byty. Z kolei niektórzy filozofowie nauki wcale nie są przekonani co do tego, że teorie naukowe stosują się do rzeczywistości w taki sposób, że są one w stanie np. wyjaśnić wszystkie parametry ruchu (np. Cartwright 1983; Dupré 1993; choć por. Nowak 1977; 1980).

⁵ Należy zaznaczyć, że Lowe nie wymienia tych tez *explicitie*. Wprowadza on je jednocześnie, swobodnie przechodząc od tez dotyczących wyjaśnień do tez dotyczących przyczyn i odwrotnie.

związaną w szczególności z tym, jakimi zasobami pojęciowymi dysponujemy, wyjaśniając takie zjawisko na poziomie psychologii ludowej. (*Nota bene* z punktu widzenia tych zasobów jesteśmy zresztą równie bezradni, jeżeli chodzi o detaliczny opis samego uniesienia ręki). Przeniesienie obserwacji dotyczących wyjaśnień na tezy ontologiczne należałoby jakoś umotywić. Wydaje się, że argument z implikacji kontrfaktycznych mógłby stanowić właśnie taką motywację.

Po drugie jednak Lowe aspiruje do tego, by jego koncepcja niekartezjańskiego dualizmu substancjalnego była zgodna z wiedzą neuronaukową.⁶ Otóż nie wydaje się, aby twierdzenia (N-) czy (ON-) były z nią zgodne. W sprzeczności z nią stoi twierdzenie, zgodnie z którym zdarzenia neuronalne nie powodują, że do uniesienia ręki w ogóle doszło — przecież gdyby odpowiednie neurony nie były aktywne w odpowiedni sposób, to nie kurczyłyby się odpowiednie mięśnie i tym samym ręka by się nie uniosła.

Lowe formułuje swoje stanowisko w następujący sposób:

Różnica między tymi dwoma rodzajami wyjaśnień przyczynowych uwidacznia się jasno [...], gdy rozważymy ich odpowiednie implikacje *kontrfaktyczne*. [1] Gdyby sprawca nie zdecydował się unieść ręki, to uniesienie ręki w *ogóle* by nie nastąpiło [...]. [2] Nie wydaje się jednak, aby można było określić jakiegokolwiek zdarzenie neuronalne bądź zbiór zdarzeń neuronalnych, których niewystąpienie miałyby *dokładnie takie same* konsekwencje, jak niewystąpienie decyzji sprawcy. [3] Jedyne co można powiedzieć to to, że jeżeli to czy tamto zdarzenie neuronalne, bądź zbiór zdarzeń neuronalnych by nie wystąpiło, to uniesienie ręki dokonałoby się w nieco inny sposób — może byłoby gwałtowniejsze czy też szybsze — ale *nie* że ręka pozostałaby w spoczynku bądź też poruszyłaby się w zupełnie inny sposób (Lowe 2008, s. 103).

Twierdzenie oznaczone jako [1] jest twierdzeniem psychologii ludowej i nie będziemy go podważać (jest to bowiem twierdzenie przyjmowane przez tych fizykalistów, w których wymierzona jest argumentacja Lowe'a). Twierdzenie [2] jest dyskusyjne, ale istnieją pewne racje, by je utrzymywać. Prawie na pewno nie jest tak, że mianem „intencji, by podnieść rękę” (nie mówiąc o intencjach bardziej złożonych) można określić aktywność jednego neuronu. Ponadto jeżeli paradygmat konekcjonistyczny daje nam lepszy wgląd w zrozumienie realizacji funkcji poznawczych przez sieci neuronowe, to można mieć zasadne wątpliwości, czy dawałoby się wykroić jakąkolwiek grupę neuronów, której aktywność można byłoby określić jako realizowanie pewnej intencji.⁷ Zupełną zagadką pozostaje jednak to, jakie twierdzenia empiryczne

⁶ Lowe stara się wykazać (por. szczególnie 2008, rozdział 3), że jego koncepcja jest zgodna z wiedzą neuronaukową, konstruując przyczynowość mentalną w taki sposób, że jest ona zasadniczo niedostrzegalna z perspektywy neuronaukowej. Argumentację tę trzeba byłoby zbadać bliżej, czego się tutaj nie podejmę. Warto jedynie zwrócić uwagę na fakt, że nosi ona znamiona argumentacji sceptycznej. W tym samym sensie, w którym koncepcja Lowe'a jest zgodna z wiedzą naukową, zgodne z wiedzą naukową jest np. twierdzenie, że wszystkie zdarzenia w świecie mają dodatkową — za każdym razem odmienną — przyczynę supranaturalną. Zwykle mówiąc o „zgodności z wiedzą naukową”, używamy silniejszego pojęcia.

⁷ Tego rodzaju refleksja jest pożywką dla eliminatystów, por. szczególnie Ramsey *et al.* (1995).

bądź quasi-empiryczne mogłyby choćby zbliżyć nas do uzasadnienia tezy [3]. Jak zobaczymy, podstawową motywacją dla przyjęcia tej tezy jest argument z implikacji kontrfaktycznych, zgodnie z którym niewystąpienie zdarzenia mentalnego doprowadziłoby do niepodniesienia ręki, natomiast niewystąpienie zdarzenia neuronalnego nadal doprowadziłoby do podniesienia ręki, choć w nieco inny sposób.

Lowe mógłby też odwołać się do wprowadzonego przezeń pojęcia przyczynowości koincydentalnej. Przyczyny pewnego zdarzenia *S* są koincydentalne wówczas, gdy powodują *S* niezależnie od siebie, a precyzyjniej: gdy nie mają wspólnej przyczyny (Lowe 2008, s. 27-28). Według Lowe'a wyjaśnienia neuronalne odwołują się właśnie do przyczyn koincydentalnych. Otóż w tym samym sensie koincydentalna jest mikroprzyczynowość w ogóle. Jeżeli nadeptujemy na śnieg (przy temperaturze ok. -1°C), to zostawimy po sobie mokry ślad. Wyjaśnienie jest proste: nadeptanie spowodowało wzrost ciśnienia, a ten — wzrost ciepła, który doprowadził do zmiany stanu wody ze stałego w ciekły. Oczywiście możemy to, co się dzieje, wyjaśnić też na poziomie mikrozjawisk. Wzrost ciśnienia wiąże się ze wzrostem pędu, a ten ze wzrostem energii kinetycznej cząsteczek wody itd. Stosując kategorie Lowe'a, musielibyśmy powiedzieć, że być może takie mikrowyjaśnienie pozwala na zrozumienie parametrów ruchu poszczególnych cząsteczek wody, natomiast zagadką pozostaje, dlaczego do wzrostu ciepła czy do roztopienia śniegu w ogóle doszło. Odwołanie do ruchów cząsteczek jest w ten sam sposób odwołaniem do przyczyn koincydentalnych, co odwołanie do aktywności neuronów.

Jedynym uzasadnieniem Lowe'a dla tez (N-) i (ON-) pozostaje odwołanie do argumentu z implikacji kontrfaktycznych. Argument ten odgrywa zatem istotną rolę w Lowe'a koncepcji niekartezjańskiego dualizmu substancji. Ma on pokazać, że substancja mentalna nie jest tylko „cieniem” substancji fizycznej, ale posiada niezależne moce przyczynowe (Lowe 2008, s. 99).

Warto jednak podkreślić, że argument Lowe'a jest w sposób istotny oparty na pewnym konkretnym sposobie rozumienia zdań kontrfaktycznych, a mianowicie na semantyce światów możliwych. *Prima facie* nie jest zatem wcale pewne, że ustalenia Lowe'a mogłyby być powielone przy alternatywnych ujęciach zdań kontrfaktycznych. Co więcej, sam Lowe zgłaszał niezależnie szereg wątpliwości wobec semantyki światów możliwych (por. Lowe 1995).⁸ Co prawda stwierdza on *explicite*, że jego wątpliwości nie podważają argumentu z implikacji kontrfaktycznych, to jednak taki zabieg wydaje się dość osobliwy. Przerysowując nieco — jest to trochę tak, jak gdyby zwolennik teorii oksydacyjnej odwoływał się do teorii flogistonowej, by uzasadnić któreś ze swoich twierdzeń.

⁸ Na gruncie rodzimym innego typu wątpliwości zgłasza np. Tałasiewicz (2008).

2. ARGUMENT Z IMPLIKACJI KONTRFAKTYCZNYCH

Według zwolenników teorii identyczności (zarówno jednostkowej, jak i rodzajowej), zdarzenie (egzemplarz) D (decyzja podmiotu, by podnieść rękę) jest tożsame z pewnym złożonym zdarzeniem (egzemplarzem) neuronalnym N .⁹ Argumentacja Lowe'a (2006; 2008) ma wykazać prawdziwość następujących zdań kontrfaktycznych:

- (L1) „Gdyby zdarzenie N nie wystąpiło, to ręka podmiotu uniosłaby się prawie dokładnie tak samo, jak to rzeczywiście miało miejsce” (Lowe 2008, s. 105, podkreślenie autora)
- (L2) „Gdyby zdarzenie D nie wystąpiło [...], to ręka podmiotu *w ogóle by się nie uniosła*.” (Lowe 2008, s. 105, podkreślenie autora).

Jeżeli (L1) i (L2) są prawdziwe, to zdarzenia D i N nie mogą być tożsame. Lowe w ten sposób przedstawiłby potężny argument przeciwko teoriom identyczności (w tym przeciwko teoriom identyczności jednostkowej¹⁰). Argumentacja na rzecz (L1) jest niezależna od argumentacji na rzecz (L2). Kolejno zrekonstruujemy obie, rozpoczynając od dowodu na (L1).

Lowe przyjmuje, że zachodzi relacja przyczynowa między pewnym złożonym zdarzeniem neuronalnym N a pewnym zdarzeniem behawioralnym B (tu: uniesieniem ręki); w obu wypadkach mówimy o zdarzeniach-egzemplarzach. Prawdziwe jest zatem następujące zdanie kontrfaktyczne:

- (L1a) Gdyby zdarzenie neuronalne N nie wystąpiło, to nie wystąpiłoby B .

Według standardowej interpretacji zdań kontrfaktycznych (L1a) jest prawdziwe zawsze i tylko wtedy, gdy we wszystkich najbliższych światach możliwych, gdzie N nie występuje, nie występuje też B . Aby dowieść prawdziwości (L1), Lowe musiałby pokazać, że we wszystkich tych światach występuje zdarzenie (egzemplarz) podobne do B , tj. również będące uniesieniem ręki, choć charakteryzujące się nieznacznie odmiennymi parametrami.

Lowe uważa, że może tego dowieść, odwołując się do relacji bliskości między światami narzucanej przez standardową interpretację zdań kontrfaktycznych. Twierdzi, że w najbliższych światach możliwych, gdzie nie występuje N , wystąpi inne zdarzenie neuronalne N^* , które musi być *bardzo* podobne do zdarzenia rzeczywistego N . Podobieństwo jest wymuszone przez relację bliskości:

⁹ Jak zwraca uwagę recenzent, Lowe konstruuje decyzje jako fakty raczej niż zdarzenia. Tak jest istotnie. Ponieważ jednak argumentacja Lowe'a jest wymierzona w tezę identyczności jednostkowej, więc Lowe w kontekście tej argumentacji *de facto* mówi o zdarzeniach-egzemplarzach.

¹⁰ W książce (2008) Lowe rozszerza swoje wnioski również na wersje fizykalizmu odrzucające nawet teorię identyczności jednostkowej.

[...] jeżeli N jest [...] sumą tej ogromnej liczby pojedynczych zdarzeń neuronalnych, to *najbliższym* światem, w którym samo N nie występuje, jest świat, w którym występuje *inne* niezwykle złożone zdarzenie neuronalne N^* , różniące się *tylko odrobinę* od N pod względem pojedynczych zdarzeń, których jest sumą. [...] Dowolny świat możliwy, w którym występuje zdarzenie neuronalne różniące się od N *bardziej* niż w ten minimalny sposób po prostu nie kwalifikuje się jako *najbliższy* świat możliwy, w którym N nie występuje (Lowe 2008, s. 105).

Jeżeli przyjąć takie rozumienie tego, co wymaga relacja bliskości, pozostałe kroki nie budzą większych zastrzeżeń. Ponieważ N różni się od N^* tylko aktywnością co najwyżej paru neuronów, to — przy założeniu masywnego rozproszenia aktywności neuronalnej¹¹ — N^* będzie się charakteryzowało podobnymi (do N) mocami przyczynowymi, a w związku z tym również doprowadzi do uniesienia ręki B^* , które jedynie nieznacznie różnić się będzie od rzeczywistego uniesienia ręki B . Lowe wyprowadza stąd wniosek, że gdyby zdarzenie neuronalne N , które spowodowało uniesienie ręki B , nie wystąpiło, to ręka tak czy owak by się uniosła — choć jej uniesienie mogłoby charakteryzować nieco inne tempo, trajektoria, czas, etc.

Argumentację na rzecz (L1) można w skrócie przedstawić w następujący sposób:

(L1a) Gdyby N nie wystąpiło, to nie wystąpiłoby B . (Założenie fizykałisty).

W najbliższych światach możliwych, gdzie N nie występuje:

- (b) występuje N^* , gdzie N^* jest bardzo podobne do N (wymóg bliskości),
- (c) moce przyczynowe N^* są bardzo podobne do mocy przyczynowych N (z (b)),
- (d) N^* powoduje B^* (z (c)),
- (e) B^* jest bardzo podobne do B (z (c), (d)).

(L1) Gdyby zdarzenie N nie wystąpiło, to ręka podmiotu uniosłaby się w prawie dokładnie taki sam sposób, jak to rzeczywiście miało miejsce (z (b)-(e)).

Przejdźmy do argumentacji Lowe'a na rzecz (L2). Jest ona w zasadzie przedstawiona w następującym fragmencie:

Gdyby D nie wystąpiło — gdyby sprawca nie dokonał właśnie tego aktu wyboru, którego dokonał, aby unieść rękę — to jego ręka *w ogóle by się nie uniosła*. Byłoby — jak sądzę — niedorzecznością przypuszczać, że gdyby sprawca nie podjął właśnie *tej* konkretnej decyzji D , to podjąłby decyzję w zasadzie nieodróżnialną od D — innymi słowy, *inną* decyzję, by unieść rękę w ten sam sposób lub w prawie ten sam sposób. Przeciwnie, gdyby sprawca nie podjął *tej* decyzji, to podjąłby albo zupełnie inną decyzję, albo nie podjąłby żadnej decyzji. W każdym

¹¹ Można oczywiście powątpiewać, czy zawsze będzie tak, że odmienna aktywizacja jednego czy dwóch neuronów nie doprowadzi do poważniejszych niż minimalne zmian. (Dziękuję p. Kindze Jęczyńskiej za zwrócenie na to uwagi.) Są to uzasadnione wątpliwości empiryczne, które jednak odsuwamy na bok z uwagi na fakt, że argumentacja Lowe'a jest błędna w sposób bardziej zasadniczy i niezależny od rozstrzygnięć kwestii empirycznych.

wypadku [...] jego ręka *nie uniosłaby* się prawie dokładnie tak samo, jak to rzeczywiście miało miejsce (Lowe 2008, s. 105).

Argumentację na rzecz (L2) przedstawić można w następujący sposób:

- (L2a) Gdyby żadne *D*-podobne zdarzenie nie wystąpiło w *t*, to żadne *B*-podobne zdarzenie nie wystąpiłoby w *t'*. (Teza psychologii ludowej)
- (L2b) Nieprawda, że, gdyby *D* nie wystąpiło w *t*, to inne *D*-podobne zdarzenie wystąpiłoby w *t*. (Argument z ziarnistości indywiduacji)
- (L2c) Gdyby *D* nie wystąpiło w *t*, to żadne inne *D*-podobne zdarzenie nie wystąpiłoby w *t*. (Teza Lowe'a)
- (L2d) Gdyby *D* nie wystąpiło w *t*, to żadne *B*-podobne zdarzenie nie wystąpiłoby w *t'*. (z (L2a), (L2c))
- (L2) „Gdyby *D* nie wystąpiło w *t*, to ręka w ogóle by się nie uniosła.” (Lowe 2008, s. 105) (L2d)

(L2a) jest tezą psychologii ludowej, przyjmowaną przez fizykalistów, w których wymierzony jest ten argument, więc nie należy jej tu kwestionować. Teza (L2b) wydaje się zasadna, a Lowe popiera ją dodatkowo ciekawą argumentacją, powołując się na odmienną ziarnistość kryteriów indywiduacji dla decyzji (odwołujących się do treści) oraz kryteriów indywiduacji dla ruchów ciała (odwołujących się np. do dokładnej trajektorii czy innych parametrów fizycznych ruchu). Lowe posługuje się wcześniejszym ustaleniem (L1), a mianowicie twierdzeniem, że w każdym świecie możliwym, w którym nie występuje *D*, wystąpi też ruch ręki podobny do *B*. Twierdzenie, że w każdym świecie możliwym, w którym nie występuje *D*, występuje inne zdarzenie *D'* (różne od *D*), które prowadzi do ruchu ręki *B'*, wiązałoby się z przekonaniem, że drobiazgowość kryteriów indywiduacji dla ruchów ciała musi być zbliżona do drobiazgowości kryteriów indywiduacji dla decyzji, co wydaje się wątpliwe z psychologicznego punktu widzenia. Nigdy nie podejmujemy decyzji, aby unieść rękę zgodnie z precyzyjnie — z fizycznego punktu widzenia — określoną trajektorią ruchu. Jeżeli jednak miałyby tak być, to — ujmując tę argumentację obrazowo — różnych decyzji mogłoby nie starczyć, by „pokryć” różne ruchy ręki.

Przesłanką budzącą poważne wątpliwości jest jednak (L2c). Jeżeli przyjmiemy (L2c), to reszta argumentacji da się przeprowadzić. W §5 pokażę, że dysponujemy wystarczającymi racjami, by (L2c) odrzucić.

Gdyby argumentacja Lowe'a na rzecz tez (L1) i (L2) była poprawna formalnie i materialnie, to należałoby odrzucić tezę o identyczności zdarzeń (egzemplarzy) *N* i *D*. Jest to teza przyjmowana zarówno na gruncie teorii identyczności rodzajowej, jak i teorii identyczności jednostkowej. Co więcej, przyjęcie (L1) i (L2) uzasadnia

dualistyczną koncepcję przyczynowości Lowe'a. Różnice w implikacjach kontrfaktycznych niewystąpienia z jednej strony złożonego zdarzenia neuronalnego, a z drugiej strony intencji mają pokazać fundamentalną rolę przyczynową odgrywaną przez intencję, a nieodgrywaną przez złożone zdarzenie neuronalne. Tezy (L1) i (L2) miałyby nas skłaniać do poglądu, że to właśnie podmiot podejmujący decyzję jest przyczynowo odpowiedzialny za to, że ręka sprawcy się w ogóle porusza, a nie — jak sądzą fizykaliści — jakies zdarzenie neuronalne. Zdarzenie neuronalne odpowiada jedynie za detale tego ruchu. Gdyby tak było, to rewizji podlegałoby szereg przekonań, które żywimy.

Przejdźmy jednak do pokazania, że argumentacja Lowe'a jest nieskuteczna. Rozpoczniemy od przytoczenia kontrprzykładów.

3. KONTRPRZYKŁADY

Należy zaznaczyć, że można żywić wątpliwości co do zdań kontrfaktycznych o kształcie (L1) czy (L2), w których poprzednik dotyczy zdarzenia (egzemplarza), a następnik — zdarzenia (typu). Podstawowy problem polega na tym, że pozwalając sobie na mówienie o zdarzeniu (typie) w następniku, ukradkiem traktujemy zdarzenie, o którym mowa w poprzedniku, również jako zdarzenie (typ), choć udajemy, że tego nie robimy. Pominiemy jednak te wątpliwości i podążymy tropem myślenia Lowe'a na nieco innych przykładach.

3.1. Przykład 1

Wyobraźmy sobie, że Jan oparzył się gorącą wodą. Uzasadnione wydaje się zatem twierdzenie, że gdyby nie chlupnęła na niego gorąca woda, to by się nie oparzył. Jednocześnie sądzymy, że zachodzi identyczność między zdarzeniem (egzemplarzem) będącym chlupnięciem gorącej wody a złożonym zdarzeniem (egzemplarzem) będącym przemieszczeniem się ogromnej liczby cząsteczek H_2O , charakteryzujących się określoną energią kinetyczną.¹² Dla uproszczenia możemy przyjąć, że w momencie, w którym woda zetknęła się ze skórą Jana, temperatura wody wynosiła 99,4°C.

Podobnie jak w argumentacji Lowe'a możemy wyróżnić trzy zdarzenia (egzemplarze):

G — chlupnięcie gorącej wody (w tym momencie, w tym miejscu, w taki sposób),

¹² Oczywiście w omawianym przypadku mamy też do czynienia z identycznością typów — temperatura cieczy jest średnią energią kinetyczną jej cząsteczek. Jednakże utożsamienie zdarzenia (typu) chlupnięcia gorącej wody ze zdarzeniem (typem) ruchem ogromnej liczby cząsteczek H_2O , charakteryzujących się taką a taką energią kinetyczną, może już nastęrczać dodatkowych trudności. Można przypuszczać, że byłyby one zbieżne z trudnościami związanymi z utożsamieniem zdarzenia (typu) decyzji, by podnieść rękę, ze zdarzeniem (typem) aktywnością pewnych neuronów.

T — zdarzenie złożone z ruchów cząsteczek H_2O , charakteryzujących się taką i taką energią kinetyczną (w tym momencie, w tym miejscu, w taki sposób),

O — oparzenie Jana (w tym momencie, w tym miejscu, w taki sposób).

Wykażemy, że możemy użyć formy argumentacji Lowe'a, aby dowieść fałszywego twierdzenia:

(C1) Gdyby T nie wystąpiło, to tak czy owak doszłoby do oparzenia Jana w prawie identyczny sposób, w jaki to się rzeczywiście stało.

Zacznijmy od zdania kontrfaktycznego, które po prostu oddaje zajęcie relacji przyczynowej między złożonym mikrozdaniem T a oparzeniem O :

(C1a) Gdyby T nie wystąpiło, to nie wystąpiłoby O .

Teraz pytamy za Lowe'em, co *by się stało*, gdyby T nie wystąpiło. Innymi słowy, rozważamy zbiór światów możliwych najbliższych rzeczywistości, gdzie T nie występuje. Przyjmując Lowe'a rozumienie bliskości światów, w światach najbliższych rzeczywistości, gdzie nie wystąpi T , wystąpi inne złożone mikrozdanie T^* , polegające na analogicznym ruchu cząsteczek H_2O , które różnić się będzie tylko tym, że jedna czy dwie cząsteczki H_2O charakteryzować się będą inną energią kinetyczną (być może temperatura zmieni się nieznacznie, ale będzie to różnica w zasadzie niewykrywalna w pomiarach). To znaczy, że w świecie tym Jan również zostanie oparzony (wystąpi zdarzenie O^*), które będzie się różniło nieznacznie tylko od O . W ten sposób dowodzimy (C1).

Wydaje się, że mamy podobne podstawy, by przyjąć:

(C2) Gdyby Jan nie został ochlpany gorącą wodą (nie wystąpiłoby G), to Jan nie oparzyłby się (nie wystąpiłoby zdarzenie O -podobne),

jakie mamy w przypadku (L2). Możemy zastosować analogiczną argumentację:

(C2a) Gdyby żadne G -podobne zdarzenie nie wystąpiło w t , to żadne O -podobne zdarzenie nie wystąpiłoby w t' .

(C2b) Nieprawda, że gdyby G nie wystąpiło w t , to inne G -podobne zdarzenie wystąpiłoby w t .

(C2c) Gdyby G nie wystąpiło w t , to żadne inne G -podobne zdarzenie nie wystąpiłoby w t .

(C2d) Gdyby G nie wystąpiło w t , to żadne O -podobne zdarzenie nie wystąpiłoby w t' .

(C2) Gdyby G nie wystąpiło w t , to Jan by się nie oparzył w t' .

(C2a) jest uzasadnione, o ile uzasadnione jest zdroworoządkowe przekonanie, że w normalnych warunkach chlupnięcie odpowiedniej ilości gorącej wody powoduje

oparzenie osoby, która w ten sposób zostaje ochlapana. Wydaje się też, że można powielić argumentację Lowe'a z odmiennej ziarnistości kryteriów indywiduacji, by uzasadnić (C2b). Przecież kryteria indywiduacji dla ochlapań gorącą wodą są o wiele mniej drobiazgowo niż kryteria indywiduacji dla ruchów cząsteczek H_2O . (C2c) jest problematyczne podobnie jak (L2c); sądząc, że mamy podobne podstawy, by przyjąć (C2c) co (L2c). (C2d) jest już tylko wyciągnięciem wniosku z (C2a) i (C2c), a (C2) — zreformulowaniem (C2d).

Jeżeli tak, to udowodniłiśmy tym samym, że zdarzenie (egzemplarz) G (chlupnięcie gorącej wody) nie jest tożsame ze zdarzeniem (egzemplarzem) T (przemieszczeniem się ogromnej liczby cząsteczek H_2O , charakteryzujących się taką a taką energią kinetyczną). Co więcej (C1) i (C2) dają podstawy, by twierdzić, że mamy do czynienia z dwoma rodzajami przyczynowości: makro i mikro. Chlupnięcie gorącą wodą determinuje to, że do oparzenia w ogóle doszło, natomiast nie determinuje detalicznych parametrów tego oparzenia („makroprzyczynowość”). Natomiast przemieszczenie ogromnej liczby cząsteczek H_2O determinuje detaliczne cechy oparzenia — jego dokładny kształt, stopień etc., natomiast nie determinuje tego, że do oparzenia w ogóle doszło („mikroprzyczynowość”).

Wnioski te wydają się nie do zaakceptowania. Coś jest nie tak. (C1) jest po prostu fałszywe. Nie jest prawdą, że gdyby nie wystąpiło złożone mikro zdarzenie T (polegające na przemieszczeniu się ogromnej liczby cząsteczek H_2O), to Jan i tak zostałby oparzony. Nie mamy powodu, aby przypuszczać, że gdyby nie wystąpiło mikro zdarzenie T , polegające na przemieszczeniu się ogromnej liczby cząsteczek H_2O , to wystąpiłoby inne zdarzenie T^* , polegające na przemieszczeniu się tych samych cząsteczek, z których jedna czy dwie różniłyby się pod względem ich energii kinetycznej. Zdarzenie T^* mogłoby wystąpić, ale równie dobrze energia kinetyczna większości tych cząsteczek mogłaby być inna od rzeczywistej.

Teza (C2) jest również fałszywa. Gdyby konkretne chlupnięcie gorącej wody G nie wystąpiło w t , to nie jest prawdą, że Jan by się w ogóle nie oparzył w t' . Przecież mogło w t wystąpić inne chlupnięcie na przykład nieco bardziej gorącej wody, w czego rezultacie Jan też zostałby oparzony w t' .

3.2. Przykład 2

W wyborach prezydenckich 4 lipca 2010 roku Polacy oddali 53% głosów na Bronisława Komorowskiego, a 47% na Jarosława Kaczyńskiego. W rezultacie Polacy wybrali Bronisława Komorowskiego na Prezydenta RP. Bronisław Komorowski został wybrany Prezydentem RP zgodnie z wolą społeczeństwa polskiego.

Wyróżnijmy trzy zdarzenia (egzemplarze):

L — głosowanie Polaków (4 lipca 2010)

P — wybór Bronisława Komorowskiego na Prezydenta RP (7 lipca 2010 roku)

W — wyrażenie woli społeczeństwa polskiego, by prezydentem RP został Bronisław Komorowski (4 lipca 2010).

Argumentacja Lowe'a pozwala dowieść, że

(W1) Gdyby *L* nie wystąpiło, to tak czy owak doszłoby do wyboru Bronisława Komorowskiego na Prezydenta RP.

Zaczynamy od niewinnego zdania kontrfaktycznego, które po prostu oddaje zajście relacji przyczynowej między złożonym zdarzeniem *L* a wyborem Prezydenta RP:

Gdyby *L* nie wystąpiło, to nie wystąpiłoby *P*.

Zdanie jest niekontrowersyjne, gdyż mówimy o zdarzeniach-egzemplarzach. Wymóg bliskości nakazuje, aby dobrać światy najbliższe w taki sposób, by tylko parę z głosujących osób głosowało inaczej niż to miało w rzeczywistości miejsce (*L**). Inne wybory paru osób nie wpłyną na wynik głosowania, a zatem tak czy owak Bronisław Komorowski będzie wybrany na Prezydenta RP. Wykazaliśmy zatem, że we wszystkich najbliższych światach możliwych, gdzie nie wystąpiło *L*, wystąpiło zdarzenie *P*-podobne, czyli wykazaliśmy, że (W1) jest prawdziwe.

Moglibyśmy też zastosować wnioskowanie Lowe'a, by ustalić, że

(W2) Gdyby *W* nie wystąpiło, to nie doszłoby do wyboru Bronisława Komorowskiego na Prezydenta RP.

Wnioskowanie na rzecz (W2) przebiegałoby podobnie jak wyżej. Przyjmujemy zdroworozsądkowe przekonanie (W2a), że gdyby społeczeństwo polskie nie wyraziło woli, by prezydentem był Bronisław Komorowski (nie wystąpiłoby zdarzenie *W*-podobne), to Bronisław Komorowski nie zostałby prezydentem (nie wystąpiłoby zdarzenie *P*-podobne). Argument z odmiennej ziarnistości kryteriów indywidualacji dla wyrażen woli społecznej z jednej strony a głosowań z drugiej pozwala na przyjęcie (W2b): nie jest prawdą, że gdyby *W* nie wystąpiło, to wystąpiłoby inne zdarzenie *W*-podobne, tj. doszłoby do innego wyrażenia woli wyboru Bronisława Komorowskiego na Prezydenta RP. Przyjmujemy za Lowe'em najbardziej problematyczną przesłankę (W2c): gdyby *W* nie wystąpiło, to nie wystąpiłoby zdarzenie *P*-podobne, tj. nie doszłoby do wyboru Bronisława Komorowskiego na Prezydenta RP.

W konsekwencji wykazalibyśmy, że zdarzenie jednostkowe (z dnia 4 lipca 2010), będące wyrażeniem woli społeczeństwa polskiego, aby Bronisław Komorowski został Prezydentem RP, różni się od zdarzenia jednostkowego (z dnia 4 lipca 2010) będącego głosowaniem Polaków w wyborach prezydenckich. Co więcej, można przypuszczać, że mamy tu do czynienia z dwoma rodzajami przyczynowości — „indywidualową” i „zespołową”. Wyrażenie woli społeczeństwa polskiego zdeterminowało to, że do wyboru Komorowskiego na Prezydenta RP w ogóle doszło (przyczynowość zespołowa), natomiast głosowanie Polaków zdecydowało już tylko, w jaki sposób (np. dokładnie, z jaką przewagą) Komorowski został wybrany (przyczynowość indywidualowa).

* * *

Jestem zdania, że oba przykłady pokazują, iż argumentacja, na której opiera się Lowe, jest błędna. Można się o tym jednak przekonać, nawet jeżeli z jakiegoś powodu zakwestionuje się te przykłady jako kontrprzykłady.

4. O ARGUMENTACJI NA RZECZ TEZY (L1): STARY PROBLEM W NOWEJ ODSŁONIE

Uderzające jest to, że jedynym argumentem Lowe'a na rzecz prawdziwości tezy (L1) jest zastosowanie semantyki światów możliwych (Lewis 1973).¹³ Okazuje się jednak, że w wypadku zdań o strukturze podobnej do (L1) (por. §4.1. Struktura logiczna tezy (L1)) semantyka światów możliwych zakłada warunki prawdziwości odmienne od tych, które bylibyśmy skłonni tym zdaniom przypisać (§4.2.). Problem ten wydaje się szczególnym przypadkiem problemu poprzedników o kształcie alternatywy (§4.3. Problem poprzedników). Jeżeli jednak rzeczywiście tak jest, to nie należy stosować semantyki światów możliwych do ustalania wartości logicznej zdań typu (L1) (§4.4.).

4.1. Struktura logiczna tezy (L1)

Teza (L1) odnosi się do bardzo złożonego zdarzenia N . Warto ją tak przeformułować, by *explicite* oddać tę złożoność. Lowe mówi, że N jest „sumą [...] ogromnej liczby pojedynczych zdarzeń neuronalnych” (Lowe 2008, s. 105). Przyjmijmy, że te pojedyncze zdarzenia neuronalne to jednostkowe zdarzenia N_1, \dots, N_k polegające na takiej a nie innej aktywności neuronów n_1, \dots, n_k . Przy tym założeniu konieczne jest, by N występowało zawsze i tylko wtedy, gdy występuje zarówno N_1, \dots , jak i N_k . W takim razie teza (L1) wydaje się równoważna tezie (L1'):

(L1') Gdyby nie wszystkie zdarzenia N_1, \dots, N_k wystąpiły (gdyby nie wszystkie neurony zachowywały się tak, jak się rzeczywiście zachowują), to ręka podmiotu uniosłaby się w prawie dokładnie taki sam sposób, jak to rzeczywiście miało miejsce.

(L1') można podeprzeć analogicznym rozumowaniem. Najbliższe światy możliwe, w których spełniony jest poprzednik, tj. gdzie nie wszystkie z tych neuronów zachowują się tak, jak się rzeczywiście zachowują, to światy, w których *prawie* wszystkie z tych neuronów zachowują się tak, jak się rzeczywiście zachowują. Można zasadnie

¹³ Sam Lowe (1995) wyraża wiele wątpliwości dotyczących interpretacji zdań kontrfaktycznych, zwracając uwagę na to, że w wypadku pewnych zdań to intencja mówcy określa, które światy należy uznać za bliskie świata rzeczywistego. Jednakże stwierdza też (2008, s. 103), że wątpliwości te nie mają znaczenia dla omawianej argumentacji.

sądzić, że odmienna aktywność jednego czy dwóch z tych neuronów nie wpłynie zasadniczo na to, czy ręka się uniesie, a co najwyżej sprawi, że uniesienie ręki będzie się charakteryzowało nieco innymi parametrami.

4.2. Zdania kontrfaktyczne z zanegowaną koniunkcją w poprzedniku a semantyka światów możliwych

Warto jednak zwrócić uwagę, że zastosowanie semantyki światów możliwych do zdań kontrfaktycznych, w których poprzedniku znajduje się negacja koniunkcji, systematycznie będzie przynosić nieintuicyjne rezultaty.

Przykład 1. Załóżmy, że Lena ma alergię na sierść. Tak się akurat składa, że jej dwie najbliższe przyjaciółki Asia i Beata dostają w prezencie zwierzaka — Asia dostaje kota, a Beata — psa. Ponieważ Lena spędza z nimi dużo czasu, występuje u niej silna reakcja alergiczna. Co by się jednak stało, gdyby nie było prawdą, że obie przyjaciółki dostały w prezencie zwierzaka?

Założmy, że ktoś byłby skłonny odpowiedzieć, że:

(1#) Gdyby nie było prawdą, że obie Asia i Beata dostały swojego zwierzaka, to u Leny wystąpiłaby i tak silna reakcja alergiczna.

Wydaje się, że powinniśmy temu komuś przypomnieć, że zdania o kształcie „nieprawda, że zarówno p i q ” dopuszczają sytuację, w której „ani nie p , ani nie q ”. Innymi słowy pytanie dopuszcza sytuację, w której żadna z przyjaciółek nie dostaje swego czworonoga. W takim wypadku, *ceteris paribus*, reakcja alergiczna by u Leny nie wystąpiła. Na pewno mogłoby być tak, że u Leny i tak wystąpi reakcja alergiczna — w sytuacji, w której tylko jedna z koleżanek otrzymuje zwierzę. To znaczy jednak że:

(1) Gdyby nie było prawdą, że obie Asia i Beata dostały swego zwierzaka, to u Leny *mogłaby* i tak *wystąpić* silna reakcja alergiczna.

Prawdą jest też, że:

(1') Gdyby nie było prawdą, że obie Asia i Beata dostały swego zwierzaka, to u Leny *mogłaby nie wystąpić* silna reakcja alergiczna.

Zdanie (1#) jest zatem fałszywe.

Stosując semantykę światów możliwych¹⁴, otrzymamy jednak raczej twierdzenie (1#) niż (1) i (1'). *Ceteris paribus* światy, gdzie tylko jedna z przyjaciółek otrzymuje czworonoga, są bliższe światu rzeczywistego niż światy, w których żadna z przyjaciółek nie dostaje zwierzaka. Zatem w najbliższych światach możliwych, gdzie zwierzątko jest подарowane tylko jednej z przyjaciółek, u Leny pojawią się objawy alergiczne.

¹⁴ Zakładam, że stosujemy semantykę światów możliwych w sposób, w który to czyni Lowe. Dopuszczam możliwość, że można by użyć jakiegoś sposobu jej zastosowania (np. przez odpowiednie skonstruowanie funkcji selekcji światów) tak, by nie generowała podobnych rezultatów.

Werdykt semantyki światów możliwych odbiega zatem od naszych intuicji.

Przykład 2. Aby seminarium zostało uruchomione, zapisanych musi być co najmniej pięcioro studentów. W programie, w ramach którego seminarium te są proponowane, zapisanych jest dziesięciu studentów. W rzeczywistości wszystkie te osoby zapisały się na seminarium z ontologii, w związku z czym zostało ono uruchomione, a nie zostało uruchomione seminarium konkurencyjne z epistemologii. Co by było, gdyby nie wszystkie z tych osób zapisały się na seminarium z ontologii?

Na to pytanie nie ma jednoznacznej odpowiedzi, dopóki nie wiemy, ile osób zapisało się na jedno bądź drugie seminarium. To, że nie wszystkie osoby zapisały się na seminarium z ontologii dopuszcza trzy istotnie różne sytuacje: (i) większość studentów zapisuje się na seminarium z ontologii, (ii) tyle samo (pięcioro) osób zapisuje się na jedno i drugie seminarium, (iii) większość osób zapisuje się na seminarium z epistemologii. W niektórych z tych dopuszczonych przez poprzednik możliwych sytuacji (mianowicie w (i) i w (ii)) seminarium z ontologii zostanie uruchomione. Natomiast w sytuacji (iii) seminarium z ontologii nie zostanie uruchomione. Wydaje się, że (przynajmniej bez dodatkowej wiedzy o tej sytuacji) możemy co najwyżej wnosić, że następujące zdania kontrfaktyczne są prawdziwe:

- (2) Gdyby nie wszyscy studenci zapisali się na seminarium z ontologii, to seminarium to mogłoby i tak zostać uruchomione.
- (2') Gdyby nie wszyscy studenci zapisali się na seminarium z ontologii, to seminarium to mogłoby nie zostać uruchomione.

Stosując semantykę światów możliwych w powyższy sposób, otrzymujemy jednak inny werdykt. Najbliższe światy możliwe, w których nie wszystkie dziesięć osób rejestruje się na seminarium z ontologii, to światy, w których prawie wszystkie te osoby (z wyjątkiem jednej, czy dwóch) się na nie rejestrują. W takim jednak wypadku seminarium z ontologii zostanie uruchomione, a więc mamy podstawy przyjąć:

- (2#) Gdyby nie wszyscy studenci zapisali się na seminarium z ontologii, to seminarium to zostałoby i tak uruchomione.

(2#) jest jednak fałszywe. Można z powodzeniem przyjąć twierdzenie słabsze (2), a twierdzenie (2#) odrzucić jako za silne.

4.3. Problem poprzedników o kształcie alternatywy

Powyższe zastosowanie semantyki światów możliwych do zdań kontrfaktycznych, których poprzedniki są zanegowanymi koniunkcjami, systematycznie odbiega od naszych intuicji. Wydaje się, że można to rozumieć po prostu jako jeszcze jeden przejaw starego problemu poprzedników o kształcie alternatywy. W końcu negacja koniunkcji dwóch zdań jest równoważna alternatywie negacji tych zdań.

Wcześni krytycy semantyki światów możliwych (Creary & Hill 1975; Ellis et al. 1977; Fine 1975; Nute 1975) zwracali uwagę, że nie oddaje ona dobrze naszych intuicji dotyczących zdań kontrfaktycznych, których poprzednikami są alternatywy. Rozważmy następujące zdanie:

- (6#) Gdyby w czasie II Wojny Światowej Hiszpania była sojusznikiem Ameryki lub Niemiec, to Hitler byłby zadowolony.

Uznalibyśmy to zdanie za fałszywe. Nawet jeżeli zgodzimy się, że Hitler byłby zadowolony, gdyby Hiszpania, która *de facto* zachowała neutralność, była sojusznikiem Niemiec, to trudno się zgodzić, że Hitler byłby zadowolony, gdyby była sojusznikiem Ameryki. Jednakże najbliższe światy możliwe, w których spełniona jest alternatywa „Hiszpania jest sojusznikiem Ameryki lub Niemiec”, to światy, w których Hiszpania jest sojusznikiem Niemiec (Hiszpania zachowała co prawda neutralność, ale było jej dużo bliżej do państw Osi niż do Aliantów). Zatem w najbliższych światach możliwych, gdzie spełniona jest ta alternatywa, Hitler jest zadowolony. Tutaj też werdykt semantyki światów możliwych różni się od naszych intuicji.

Innym przykładem tego problemu — nawiązującym do naszego przykładu 1 — byłoby zdanie:

- (7#) Gdyby albo Asia nie dostała zwierzaka albo żadna z przyjaciółek nie dostała zwierzaka, to u Leny wystąpiłaby i tak silna reakcja alergiczna.

Zdanie to jest ewidentnie fałszywe — gdyby żadna z przyjaciółek nie dostała zwierzaka, to alergia by u Leny nie wystąpiła.

W podobny sposób na podstawie przykładu 2 można łatwo skonstruować fałszywe zdanie kontrfaktyczne, którego poprzednik jest alternatywą:

- (8#) Gdyby albo tylko jeden z dziesięciu studentów albo wszyscy studenci zapisali się na seminarium z epistemologii (raczej niż na seminarium z ontologii, na które się *de facto* zapisali), to seminarium z ontologii i tak zostałoby uruchomione.

Zdanie to jest fałszywe, ponieważ w sytuacji, gdy wszyscy studenci zapisują się na seminarium z epistemologii, seminarium z ontologii nie zostanie uruchomione. Jednakże światy najbliższe, w których spełniona jest alternatywa „albo tylko jeden z dziesięciu studentów, albo wszyscy studenci zapisali się na seminarium z epistemologii” to światy, w których tylko jeden z dziesięciu studentów zapisał się na seminarium z epistemologii. W tych światach seminarium z ontologii zostaje uruchomione.

Okazuje się, że nie ma łatwego sposobu rozwiązania tej kwestii (dobry przegląd proponowanych rozwiązań znajduje się w pracy (Nute 1984) oraz (Nute, Cross 2001)), a poszukiwania trwają nadal (por. np. Alonso-Ovalle 2009; Briggs 2012). Nie ma ustalonego rozwiązania, natomiast panuje dość powszechne przekonanie, że w przypadkach zdań kontrfaktycznych z poprzednikami o kształcie alternatywy semantyka światów możliwych generuje rezultaty niezgodne z naszą oceną tych zdań.

Zatem nawet jeżeli semantyka światów możliwych jest najlepszą spośród dostępnych semantyk zdań kontrfaktycznych, to nie powinniśmy jej stosować do oceny wartości logicznej zdań kontrfaktycznych z poprzednikami o kształcie alternatywy. W tych wypadkach często daje ona wyniki niezgodne z intuicyjnymi. Innymi słowy, semantyka światów możliwych w swej standardowej wersji jest *niewiarygodną* metodą oceny wartości logicznej zdań kontrfaktycznych, których poprzednikami są zdania o kształcie alternatywy bądź np. negacji koniunkcji.

4.4. Wnioski

Podsumujmy powyższe ustalenia. Po pierwsze przykłady zdań kontrfaktycznych z poprzednikami o kształcie alternatywy są dość powszechnie uznane za przypadki dla semantyki światów możliwych problematyczne. Po drugie wspomniany w §4.2. problem ze zdaniem kontrfaktycznym, których poprzedniki są negacjami koniunkcji, wydaje się być po prostu szczególnym przypadkiem problemu poprzedników o kształcie alternatywy (§4.3. Problem poprzedników o kształcie alternatywy).

Mamy zatem podstawy, by uznać semantykę światów możliwych (przynajmniej w jej standardowym zastosowaniu) za niewiarygodną metodę oceny wartości logicznej zdań kontrfaktycznych z poprzednikami o kształcie alternatywy czy też o kształcie negacji koniunkcji. Lowe'a teza (L1) ma właśnie taki kształt (§4.1.). Jest to zdanie kontrfaktyczne, którego poprzednikiem jest negacja wieloczłonowej koniunkcji. Ponieważ *jedynym* argumentem Lowe'a na rzecz tezy (L1) jest rozwiązanie poponowane przez semantykę światów możliwych, to powinniśmy argument Lowe'a za prawdziwością tezy (L1) zdecydowanie odrzucić. Minimalny wniosek, który powinniśmy wysunąć, to ten, że *nie mamy wystarczających podstaw*, by uznać prawdziwość tezy (L1). Argument Lowe'a za taką podstawę z powyższych względów uchodzić nie może.

Jeżeli natomiast trafne są nasze intuicje dotyczące zdań tego kształtu, to mamy podstawy, by sądzić, że teza (L1) jest nie tylko nieuzasadniona, lecz także fałszywa.¹⁵ Jeżeli ocena wartości logicznej zdania kontrfaktycznego z poprzednikiem o kształcie negacji koniunkcji *ceteris paribus* uwzględnia wszystkie możliwości realizacji tej negacji, w tym również sytuację, w której wszystkie człony koniunkcji mają inną niż rzeczywista wartość logiczną, to zastanawiając się nad tym, co by było, gdyby nie wszystkie neurony zachowywały się tak, jak się zachowują, musimy też uwzględnić możliwość, gdy żaden z tych neuronów nie zachowuje się tak, jak w świecie rzeczywistym. W tym wypadku — a także w wielu innych, gdzie aktywność wystarczająco wielu neuronów różni się od rzeczywistej — mamy podstawy, by

¹⁵ Taki werdykt wydaje się też zgodny z propozycjami przedstawionymi w pracach (Alonso-Ovalle 2009; Briggs 2012). Warto nadmienić, że Briggs (2012) *explicitie* podziela powyższe intuicje dotyczące zdań kontrfaktycznych z negacją koniunkcji w poprzedniku.

sądzić, że ręka się nie uniesie. Teza (L1) jest zatem fałszywa. Prawdziwe są natomiast zdania słabsze:

- (9) Gdyby nie wszystkie zdarzenia N_1, \dots, N_K wystąpiły (gdyby nie wszystkie neurony zachowywały się tak, jak się rzeczywiście zachowują), to ręka podmiotu *mogłaby się unieść* w prawie dokładnie taki sam sposób, jak to rzeczywiście miało miejsce.
- (10) Gdyby nie wszystkie zdarzenia N_1, \dots, N_K wystąpiły (gdyby nie wszystkie neurony zachowywały się tak, jak się rzeczywiście zachowują), to ręka podmiotu *mogłaby się* w ogóle *nie unieść*.

Analogiczne rozumowanie pozwala na odrzucenie kontrprzykładów (C1) i (W1). W świetle powyższych rozważań nie mamy wystarczających podstaw, by którąkolwiek z tych tez przyjąć. Jeżeli natomiast nasze intuicje dotyczące zdań tego kształtu są trafne, to możemy co najwyżej przyjąć analogony słabszych twierdzeń (9)-(10).

5. ARGUMENTACJA NA RZECZ (L2): 'WYSTĄPIŁOBY' A 'MOGŁOBY WYSTĄPIĆ'

Jak pamiętamy, we wnioskowaniu na rzecz (L2) wątpliwość wzbudza kluczowa przesłanka (L2c):

- (L2c) Gdyby D nie wystąpiło w t (gdyby sprawca nie podjął właśnie tej decyzji w t , aby podnieść rękę), to żadne inne D -podobne zdarzenie nie wystąpiłoby (nie podjąłby innej decyzji by podnieść rękę) w t .

Pokażemy, że jest ona fałszywa.

Przypomnijmy argumentację Lowe'a na rzecz (L2c):

Gdyby D nie wystąpiło — gdyby sprawca nie dokonał właśnie tego aktu wyboru, którego dokonał, aby unieść rękę — to jego ręka w ogóle by się nie uniosła. Byłoby — jak sądzę — niedorzecznością przypuszczać, że gdyby sprawca nie podjął właśnie *tej* konkretnej decyzji, D , to podjąłby decyzję w zasadzie nieodróżnialną od D — innymi słowy, *inną* decyzję by unieść rękę w ten sam sposób lub w prawie ten sam sposób (Lowe 2008, s. 105).

Lowe wydaje się tutaj argumentować przeciwko wyobrażonemu oponentowi, nieprzyjmującemu (L2c). Lowe wyobraża sobie, że jego adwersarz przyjmuje, że „gdyby sprawca nie podjął właśnie *tej* konkretnej decyzji, D , to *podjąłby* [moje pokreślenie] [...] *inną* decyzję, by unieść rękę w ten sam sposób lub w prawie ten sam sposób” (Lowe 2008, s. 105). Oczywiście ktoś, kto nie przyjmuje (L2c), *może* twierdzić to, co Lowe rozważa (i co odrzuca, przyjmując (L2b)), a mianowicie:

- (1) Gdyby D nie wystąpiło w t , to D -podobne zdarzenie *wystąpiłoby* w t .

Jednakże ktoś, kto nie przyjmuje (L2c), może też twierdzić:

(2) Gdyby D nie wystąpiło w t , to D -podobne zdarzenie *mogłoby* wystąpić w t .

Jasno widać, że obie tezy (1) i (2) wykluczają (L2c).

Oczywiście teza (1) jest dużo silniejsza niż (2), a zatem też dużo trudniej jest wykazać jej prawdziwość. Lowe twierdzi, że teza (1) jest fałszywa, co wyraża w przesłance (L2b), a uzasadnia argumentem z odmiennej ziarnistości kryteriów indywiduacji.

Argument z odmiennej ziarnistości kryteriów indywiduacji nie stanowi jednak uzasadnienia dla (L2c). Pamiętajmy, że Lowe rozpoczyna od założenia tezy (1), że w każdym świecie możliwym, w którym nie występuje D , występuje inne zdarzenie D' . Lowe odrzuca jednak tę tezę, ponieważ zauważa, że różnych decyzji (prowadzących do uniesienia ręki) może nie starczyć, by „pokryć” różne ruchy ręki. Jednakże ta argumentacja nie pokazuje, że fałszywe jest (2). Aby wykazać prawdziwość (2), wystarczy znaleźć *jeden* świat możliwy, w którym nie występuje D , a występuje inne zdarzenie D' , które również doprowadzi do podobnego ruchu ręki. Argumentacja z odmiennej ziarnistości kryteriów indywiduacji tu nie pomoże. Trzeba byłoby pokazać, że niemożliwa jest w tych warunkach inna niż D decyzja, której skutkiem byłby podobny ruch ręki.

Pokazanie, że (2) jest prawdziwe nie wydaje się specjalnie trudne. Przecież gdyby sprawca nie zdecydował się podnieść ręki w t , to *mógł* on zdecydować, by szybko podnieść rękę w t , albo też *mógł* on zdecydować, by rozciągnąć się przez podniesienie ręki w t etc.

Można zaoponować, że każdy, kto decyduje, by szybko podnieść rękę, decyduje jednak, by podnieść rękę, więc decyzja, by podnieść rękę szybko, wbrew pozorom nie różni się od decyzji, by podnieść rękę. Ta obiekcja upada z dwóch powodów. Po pierwsze, nie jest prawdą, że zachodzą zakładane tu relacje implikacji:

$$(\alpha \kappa\text{-uje} \rightarrow \alpha \lambda\text{-uje}) \rightarrow (\alpha \text{ ma intencję } \kappa\text{-ować} \rightarrow \alpha \text{ ma intencję } \lambda\text{-ować})$$

Ktoś może mieć intencję, by wygrać konkurs, a wcale nie mieć intencji, by albo wygrać, albo przegrać konkurs. Aktor może mieć intencję, by unieść rękę tak, by oddać cały dramat sytuacji, a wcale nie mieć intencji, by unieść rękę w jakikolwiek sposób. Po drugie, obiekcja ta myli typy i egzemplarze. Nawet gdybyśmy przyjęli, że każdy, kto decyduje, by podnieść rękę szybko, decyduje, by podnieść rękę, to implikacja ta zachodzi dla typów decyzji. Aby dowieść (2), wystarczy pokazać, że gdyby decyzja D (typu „by unieść rękę”) nie wystąpiła w t , to istnieje możliwość, że wystąpiłaby decyzja D' (innego typu „by unieść rękę szybko”). Nawet jeżeli przyjmiemy, że D' jest — i musi również być — tego samego typu co D („by unieść rękę”), to przecież wynika stąd tylko, że D i D' są tego samego typu dla pewnego typu, a mianowicie „by unieść rękę”, ale nie dla każdego typu — D' jest typu „by unieść rękę szybko”, a D — nie.

Podsumowując, (2) wydaje się zasadne, a zatem (L2c) należy odrzucić, co czyni tezę (L2) bezpodstawną.

6. O KONTRFAKTYCZNYCH IMPLIKACJACH TEZY IDENTYCZNOŚCI JEDNOSTKOWEJ

Warto w tym momencie skonstatować, jakie są implikacje kontrfaktyczne tezy identyczności jednostkowej. Wydaje się, że nie prowadzą one do żadnych kłopotów dla fizykalistów:

Gdyby *N* (neuralne zdarzenie (egzemplarz)) nie wystąpiło, to nie wystąpiłoby *B* (behawioralne zdarzenie (egzemplarz)): to konkretne uniesienie ręki). (L1a)

Gdyby *D* (mentalne zdarzenie (egzemplarz)): ta konkretna decyzja, by unieść rękę) nie wystąpiło, to nie wystąpiłoby *B*.

Oba te twierdzenia są dla fizykałisty do przyjęcia, nie prowadzą do żadnych niepożądanych implikacji.

Przytoczone argumenty mogłyby posłużyć do uzasadnienia dalszych implikacji kontrfaktycznych:

Gdyby *N* nie wystąpiło, to mogłoby wystąpić zdarzenie *B*-podobne (mogłoby dojść do uniesienia ręki).

Gdyby *N* nie wystąpiło, to mogłoby nie wystąpić zdarzenie *B*-podobne (mogłoby nie dojść do uniesienia ręki).

Gdyby *D* nie wystąpiło, to mogłoby wystąpić zdarzenie *B*-podobne (mogłoby dojść do uniesienia ręki).

Gdyby *D* nie wystąpiło, to mogłoby nie wystąpić zdarzenie *B*-podobne (mogłoby nie dojść do uniesienia ręki).

Zgodność między tymi implikacjami kontrfaktycznymi powinna utwierdzić fizykałistów w przekonaniu, że argument Lowe'a nie stanowi dla nich problemu.

7. WNIOSKI

Argument z implikacji kontrfaktycznych zawodzi. Lowe nie wykazuje prawdziwości (L1). Jak widzieliśmy, zastosowanie semantyki światów możliwych do tez takich, jak (L1) często daje rezultaty odbiegające od oczekiwanych. Nie wykazuje też prawdziwości (L2), gdyż opiera swoje wnioskowanie na fałszywej przesłance. Wykazaliśmy tym samym — wbrew temu, co twierdzi Lowe — że argument z implikacji kontrfaktycznych nie obala teorii identyczności.

Warto też podkreślić, że nieskuteczność tej argumentacji Lowe'a stanowi niezależny problem dla jego dualistycznej koncepcji przyczynowości. Widzieliśmy bowiem, że Lowe wychodzi od obserwacji dotyczących różnicy między wyjaśnieniami

intencjonalnymi i fizykalnymi działań, przenosząc te różnice na poziom ontologiczny. Otóż argument z implikacji kontrfaktycznych stanowił pewne uzasadnienie dla takiego przeniesienia. Jeżeli prawdziwe byłyby tezy (L1) i (L2), to mielibyśmy podstawy, by twierdzić, że obserwacja dotycząca wyjaśnień ma nie tylko charakter epistemologiczny, ale również ontologiczny. Niewystąpienie decyzji, by podnieść rękę, prowadziłyby do uniesienia ręki, podczas gdy niewystąpienie przyczyny neuronalnej prowadziłyby nadal do uniesienia ręki — choć w nieco inny sposób. Jak widzieliśmy, nie mamy żadnych podstaw, by przyjąć którąkolwiek z tych tez, a przeciwnie mamy podstawy, by ich nie przyjmować. Jeżeli jednak tak jest, to można mieć uzasadnione wątpliwości co do racji dla przeniesienia obserwacji, dotyczących różnic między wyjaśnieniami intencjonalnymi i neuronalnymi, na grunt ontologiczny.¹⁶

Jednocześnie powinniśmy też odrzucić tezy (N-) i (ON-). Twierdzenie, z jednej strony, że wyjaśnienia odwołujące do zdarzeń neuronalnych nie wyjaśniają, dlaczego w ogóle do ruchu ręki doszło, a z drugiej strony, że to nie zdarzenia neuronalne są przyczynami tego, że ręka w ogóle się ruszyła, wydają się *prima facie* niezgodne z wiedzą neuronaukową. O ile podstawowym uzasadnieniem, by przyjąć te tezy, jest argument z implikacji kontrfaktycznych, o tyle wykazanie jego błędności pozbawia te twierdzenia kluczowego uzasadnienia.

Warto na zakończenie skonstatować, że podstawowym problemem przedstawianej przez Lowe'a obrony stanowiska dualizmu wydaje się pewnego rodzaju argumentacyjne przestrzelenie. Lowe oczywiście twierdzi, że broni dualizmu, ale *de facto* stosuje on argumenty, których użyć można do obrony emergentystycznego pluralizmu (por. też Paprzycka 1997). Jeśli nie przyjmujemy analogicznych do Lowe'a tez dotyczących np. makrozjawisk, to nie jesteśmy skłonni zgodzić się z nim co do jego ujęcia zjawisk mentalnych.

BIBLIOGRAFIA

- Alonso-Ovalle, L. (2009), „Counterfactuals, Correlatives, and Disjunction”, *Linguistics and Philosophy* 32, 207-244.
- Briggs, R. (2012), „Interventionist Counterfactuals”, *Philosophical Studies* 160, 139-166.
- Cartwright, N. (1983), *How the Laws of Physics Lie*, Oxford: Clarendon Press.
- Creary, L. G., Hill, C. S., (1975), „Review of D. Lewis' Counterfactuals”, *Philosophy of Science* 42, 341-344.

¹⁶ Należy jednak pokreślić, że wspomniany argument nie jest jedynym sposobem uzasadnienia koncepcji ontologicznej Lowe'a. Lowe przedstawia ciekawą koncepcję miejsca dla przyczynowania mentalnego, które miałyby uzupełniać przyczynowanie fizyczne (por. w szczególności pierwsze trzy rozdziały książki (2008)). Koncepcja ta jest — jak sądzę — do podważenia przez skonstruowanie analogii podobnych do tych, które przedstawione zostały w powyższych kontrprzykładach. W niniejszym artykule nie podejmowaliśmy dyskusji z tym ogólnym poglądem. Skoncentrowaliśmy się na krytyce koncepcji Lowe'a dotyczącej, by tak rzec, podziału ról między przyczynowaniem intencjonalnym i fizycznym.

- Dupré, J. (1993), *The Disorder of Things. Metaphysical Foundations of the Disunity of Science*, Cambridge, MA: Harvard University Press.
- Ellis, B., Jackson, F., Pargetter, R. (1977), „An Objection to Possible-World Semantics for Counterfactual Logics”, *Journal of Philosophical Logic* 6, 355-357.
- Fine, K. (1975), „Critical Notice of D. Lewis' *Counterfactuals*”, *Mind* 84, 451-458.
- Lewis, D. (1973), *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lowe, E. J. (1995), „The Truth about Counterfactuals”, *Philosophical Quarterly* 45, 41-59.
- Lowe, E. J. (1996), *Subjects of Experience*, Cambridge: Cambridge University Press.
- Lowe, E. J. (2006), „Non-Cartesian Substance Dualism and the Problem of Mental Causation”, *Erkenntnis* 65, 5-23.
- Lowe, E. J. (2008), *Personal Agency. The Metaphysics of Mind and Action*, Oxford: Oxford University Press.
- Nowak, L. (1977), *Wstęp do idealizacyjnej teorii nauki*, Warszawa: PWN.
- Nowak, L. (1980), *The Structure of Idealization*, Dordrecht/Boston: Reidel.
- Nute, D. (1984), *Conditional Logic*, w: D. Gabbay, F. Guenther (red.), *Handbook of Philosophical Logic*, vol. 2, Dordrecht: Reidel, 387-439.
- Nute, D. (1975), „Counterfactuals and the Similarity of Worlds”, *Journal of Philosophy* 72, s. 773-778.
- Nute, D., Cross, C.B. (2001), „Conditional Logic”, [w:] D. Gabbay, F. Guenther (red.), *Handbook of Philosophical Logic*, vol. 4, 2nd ed., Dordrecht: Kluwer, 1-98.
- Paprzycka, K. (1997), „Review of E. Jonathan Lowe's *Subjects of Experience*”, *Philosophy in Review* 17, 45-47.
- Ramsey, W., Stich, S., Garon, J. (1995), „Connectionism, Eliminativism and the Future of Folk Psychology”, [w:] C. Macdonald i G. Macdonald (red.), *Connectionism. Debates on Psychological Explanation*, Oxford: Basil Blackwell, s. 311-338.
- Tałasiewicz, M. (2008), „Dlaczego nie lubię semantyki światów możliwych”, *Przegląd Filozoficzny* 17-2, 211-216.
- von Wright, G. H. (1971), *Explanation and Understanding*, Ithaca, NY: Cornell University Press.
- Wilson, G. M. (1989), *The Intentionality of Human Action*, Stanford: Stanford University Press.