

# Agnieszka Dębska

---

## Wnioskowanie na temat intencjonalności działania w ujęciu filozofii eksperymentalnej

---

Filozofia Nauki 21/3, 143-155

---

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](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach  
dozwolonego użytku.

Agnieszka Dębska

## **Wnioskowanie na temat intencjonalności działania w ujęciu filozofii eksperymentalnej**

Projekt filozofii eksperymentalnej, zaproponowany i zarysowany przez Knobego (Knobe 2004, Knobe, Nichols 2008), rozwija się od dziesięciu lat. Trwają dyskusje na temat przedmiotu i celów nowej dyscypliny. Według jednego ze stanowisk filozofia eksperymentalna zajmuje się rekonstruowaniem potocznych intuicji filozoficznych ludzi za pomocą badań empirycznych typowych dla nauk społecznych. Argumenty formułowane na podstawie tych danych mają służyć między innymi jako potwierdzenie pewnych tez filozoficznych, w szczególności z zakresu filozofii języka, gdzie zgodnie z anglosaską tradycją korzysta się z argumentów z intuicji kompetentnego użytkownika języka. Filozofia eksperymentalna miałaby „sprawdzać” intuicje filozofów, wyniki badań nad intuicjami grup ludzi miałyby być przeciwwagą dla tzw. „armchair philosophy” (por. Knobe 2004). Według innego, ambitniejszego stanowiska filozofia eksperymentalna rekonstruuje nie tylko potoczne intuicje dotyczące problemów filozoficznych, lecz także psychologicznie rozumiane pojęcia, które wpływają na kształt intuicji (por. np. Knobe, Nichols 2008). Zakłada się, że odpowiednio skonstruowane eksperymenty pozwolą na stwierdzenie, w jakich sytuacjach ludzie postrzegają innych jako wolnych czy intencjonalnych i co ma wpływ na tę decyzję (a pośrednio na zastosowanie danych pojęć). Tak rozumiana filozofia eksperymentalna byłaby częścią programu kognitywistycznego, którego celem jest opisanie funkcjonowania ludzkiego umysłu.

Jednym z nurtów zainteresowań filozofów eksperymentalnych jest zrekonstruowanie intuicji dotyczących „intencjonalnego działania”<sup>1</sup>. W tym wypadku punktem

---

<sup>1</sup> Ponieważ odpowiednik angielskiego wyrażenia „do something intentionally” nie występuje potocznie w języku polskim, dopuszczalnym tłumaczeniem mogłoby być: „zrobić coś umyślnie/ce-

odniesienia dla badań empirycznych były uwagi filozofów, takich jak Ron Butler (1978: 113), który opisał następujący problem:

Jeśli Brown w trakcie zwykłej gry w kości liczy na to, że wyrzuci szóstkę, i zrobi to, nie mówimy, że wyrzucił szóstkę intencjonalnie. Jeśli jednak Brown wkłada jeden nabój do pistoletu o sześciu komorach, obraca komorę, celując w Smitha, i pociąga za spust, licząc na to, że zabije Smitha, powiemy (jeśli mu się powiedzie), że zabił Smitha intencjonalnie. Jak to możliwe, skoro prawdopodobieństwo oczekiwanego rezultatu jest takie samo?

Przykładem tym Butler chciał podać w wątpliwość zdroworoządkowe przekonanie, zgodnie z którym wnioskowanie ludzi na temat intencjonalności danego czynu nie jest zależne od jego oceny moralnej ani od tego, czy działanie to jest silnie nacechowane moralnie.

Mając na uwadze tego typu argumenty, eksperymenciści zbadali empirycznie intuicje wybranej grupy osób i wywnioskowali, że rzeczywiście mają oni tendencję do oceny skutków ubocznych działania jako intencjonalnych, gdy są postrzegane jako złe, a jako nieintencjonalnych, gdy są postrzegane jako dobre (por. np. Knobe 2003, 2006). Efekt ten został nazwanym efektem Knobego lub efektem skutków ubocznych (*side-effect effect*). Został on potwierdzony w ostatnich badaniach (zob. przegląd w Knobe 2006), również międzykulturowych (Knobe, Burra 2006).

## 1. EFEKT KNOBEGO

W klasycznym eksperymencie (Knobe 2003), którego wyniki wskazują na opisaną wyżej asymetrię, uczestników podzielono na dwie grupy. Każdej z nich przedstawiono nieznacznie różniące się historyjki na temat działania pewnej firmy:

**GRUPA 1. SZKODZENIE.** Wiceprezes spółki udał się do przewodniczącego zarządu i powiedział: „Myślimy o rozpoczęciu nowego projektu. To pomoże nam zwiększyć zyski, ale także zaszkodzi środowisku”. Przewodniczący odparł: „Nie obchodzi mnie ani trochę, czy zaszkodzi to środowisku. Chcę tylko osiągnąć tak wysoki zysk, jak to możliwe. Zaczniemy nowy projekt”. Zaczęli nowy projekt. A stan środowiska rzeczywiście się pogorszył.

**GRUPA 2. POMAGANIE.** Wiceprezes spółki udał się do przewodniczącego zarządu i powiedział: „Myślimy o rozpoczęciu nowego projektu. To pomoże nam zwiększyć zyski, a także pomoże środowisku”. Przewodniczący odparł: „Nie obchodzi mnie ani trochę, czy pomoże to środowisku. Chcę tylko osiągnąć tak wysoki zysk, jak to

---

lowo/specjalnie”. Znaczenie żadnego z tych terminów nie pokrywa się jednak ze znaczeniem terminu „intentionally” w języku angielskim. Wskazuje się na dwa aspekty znaczeniowe tego wyrażenia: jeden powiązany z wiedzą o działaniu, świadomością działania („x did something knowingly”), a drugi z chęcią zrobienia czegoś („x did something willingly”). W związku ze złożonym znaczeniem terminów „intentionality” i „intentionally” w pracy zdecydowałam się na użycie w polskim przekładzie terminów „intencjonalność” i „intencjonalnie”.

możliwe. Zaczniemy nowy projekt”. Zaczęli nowy projekt. A stan środowiska rzeczywiście się polepszył.

Badanych w obu grupach zapytano: Czy prezes zarządu *zaszkodził/pomógł* środowisku intencjonalnie? 82% badanych z pierwszej grupy odpowiedziało, że prezes zarządu *zaszkodził* środowisku intencjonalnie. W drugiej grupie wyniki były odwrotne: tylko 23% badanych odpowiedziało, że prezes zarządu *pomógł* środowisku intencjonalnie.

W odpowiedzi na badania Knobego (powtórzone dodatkowo w różnych wersjach narracyjnych, w których uzyskano ten sam wynik, por. Knobe 2003) powstało wiele dalszych badań i teorii mających na celu wyjaśnienie otrzymanego efektu. Sam autor twierdził (np. Knobe 2006), że opisane zjawisko jest konsekwencją sposobu funkcjonowania naszych pojęć i sądów moralnych, które wpływają na przypisywanie (lub nieprzypisywanie) intencjonalnych działań innym ludziom. Efekt Knobego zaobserwowano nie tylko w stosunku do terminu „intencjonalne działanie”, lecz także w stosunku do takich wyrażań, jak: „desire”, „intend”, „decide”, „being in favor of”, „advocating” (Beebe, Buckwalter 2010: 480). Może to przemawiać za tym, że efekt nie jest jedynie artefaktem bazującym na wieloznaczności terminu „intentionality”.

## 2. CEL PRACY

Moim celem jest analiza efektu Knobego w szerszej perspektywie pokrewnych badań empirycznych na temat intuicji dotyczących intencjonalnego działania. W szybko powstającej literaturze przedmiotu<sup>2</sup> pojawiają się zarówno konkurencyjne wyjaśnienia efektu Knobego, jak i badania dostarczające nowych danych na ten temat. Wiele teoretycznie ciekawych interpretacji efektu Knobego nie jest w stanie wytłumaczyć wyników innych badań. Dlatego spróbuję zebrać dotychczas opublikowane wyniki badań na temat stwierdzania intencjonalności działań (nie wdając się w ich metodologiczną poprawność) i przedstawić syntezę. Wskażę najlepsze wyjaśnienie efektu Knobego, które zarazem będzie zgodne z innymi zaobserwowanymi empirycznie zjawiskami asymetrii w ocenie, czy dane działanie ma charakter intencjonalny. Przykładem takiej teorii jest tzw. teoria odpowiedzialności (dalej TO) zaproponowana przez Wright i Bengsona (2009).

Pierwszą część pracy poświęcę przedstawieniu wyników badań świadczących o asymetrii w orzekaniu, czy działanie było intencjonalne, oraz dwóm niesatysfakcjonującym wyjaśnieniom tych badań. W drugiej części przedstawiam i krytycznie analizuję teorię TO, lepiej tłumaczącą efekt Knobego i inne zjawiska asymetrii. Odwołuje się ona do kryteriów przypisywania odpowiedzialności oraz do związku między oceną, czy działanie było intencjonalne, a przypisywaniem odpowiedzialności w stosunku do działającej osoby. Ostatecznie pokazuję, że chociaż teoria odpowie-

---

<sup>2</sup> Zob. np. przegląd na stronie internetowej Knobego: <http://philosophy.yale.edu/knobe>.

działności spełnia *warunek ogólności* — to znaczy potencjalnie tłumaczy wszystkie opisane rodzaje asymetrii — to nie jest dostatecznie empirycznie potwierdzona przez dotychczasowe badania ani nie posiada jasnych kryteriów operacjonalizacji kluczowych dla niej pojęć.

### 3. ASYMETRIA W OCENIE, CZY DZIAŁANIE BYŁO INTENCJONALNE — WYBRANE WYNIKI BADAŃ I ICH INTERPRETACJE

Adekwatna teoria tłumacząca asymetryczność w ocenie intencjonalności działań powinna spełniać co najmniej trzy warunki. Po pierwsze musi uwzględniać, obok efektu Knobego, również empirycznie zbadane pokrewne zjawiska. Po drugie powinna być zgodna z wynikami dotychczasowych badań. Po trzecie powinna być przekonująca z teoretycznego i metodologicznego punktu widzenia, na przykład posiadać dobrze zdefiniowane kryteria empirycznego testowania.

Oto lista zbadanych zjawisk dotyczących twierzeń o intencjonalności działań, które powinny zostać uwzględnione w dostatecznie ogólnej teorii asymetrii:

I. Badani twierdzą zazwyczaj, że pewna czynność została wykonana intencjonalnie, gdy jej skutki są moralnie złe, a nieintencjonalnie, gdy są moralnie dobre (tzw. efekt Knobego).

II. Badani z reguły oceniają czynność przypadkową (spowodowaną np. przez brak umiejętności) lub będącą wynikiem trafu (np. w grze losowej) jako intencjonalną, kiedy jest ona nacechowana moralnie, a jako nieintencjonalną wtedy, gdy jest ona moralnie neutralna<sup>3</sup> (co odpowiada opisanemu wyżej „problemowi Butlera”).

---

<sup>3</sup> Por. np. następujące sytuacje: (A) “Jake desperately wants to win the rifle contest. He knows that he will only win the contest if he hits the bulls-eye. He raises the rifle, gets the bull’s-eye in the sights, and presses the trigger. But Jake isn’t very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild... Nonetheless, the bullet lands directly on the bull’s-eye. Jake wins the contest” i (B) “Jake desperately wants to have more money. He knows that he will inherit a lot of money when his aunt dies. One day, he sees his aunt walking by the window. He raises his rifle, gets her in the sights, and presses the trigger. But Jake isn’t very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild... Nonetheless, the bullet hits her directly in the heart. She dies instantly” (Knobe 2006). W sytuacji (A) przedstawionej jako moralnie neutralna działanie bohatera zostało ocenione jako intencjonalne przez 23% badanych. Dla porównania działanie w sytuacji moralnie nagannej (B) zostało ocenione jako intencjonalne przez 91% badanych. Także jako intencjonalne ocenione zostało analogiczne działanie, tyle że ocenione jako moralnie godne pochwały (92%): (C) “Klaus is a soldier in the German army during World War II. His regiment has been sent on a mission that he believes to be deeply immoral. He knows that many innocent people will die unless he can somehow stop the mission before it is completed. One day, it occurs to him that the best way to sabotage the mission would be to shoot a bullet into his own regiment’s communication device. [...] With that thought in mind, he raises his rifle, gets the device in his sights, and presses the trigger. But Klaus isn’t very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild... Nonetheless, the bullet lands directly in the communications device. The mission is foiled, and many innocent lives are saved” (Knobe 2006).

III. Badani mają tendencję do oceniania pewnych czynności jako intencjonalnych, gdy ich skutki uboczne są niekorzystne, ale moralnie neutralne<sup>4</sup>.

IV. Badani mogą ocenić czyjeś działania jako nieintencjonalne, nawet wtedy gdy orzekają, że były one moralnie złe i że działający ponosi za nie moralną winę (Nadelhoffer 2004)<sup>5</sup>.

#### 4. TEORIA BŁĘDU I TEORIA DECYZJI — NIEWYSTARCZAJĄCE DO WYJAŚNIENIA ASYMETRII

Niżej przedstawiam dwie teorie, które nie spełniają warunku ogólności, gdyż nie wyjaśniają wszystkich wymienionych zjawisk.

##### 4.1. Teoria błędu

Malle i Nelson (2003) wysunęli hipotezę, że w przypadku negatywnie nacechowanych działań przekonanie o moralnej winie działającego i idące za tym emocjonalne uprzedzenie do niego wpływa na ocenę intencjonalności działania. Badani winią bohatera historyjki za skutki jego decyzji i są przez to skłonni doszukiwać się informacji, które mogłyby pogorszyć jego wizerunek. Orzekają często bezpodstawnie, że złe działanie sprawcy było intencjonalne, ponieważ zadecydowali wcześniej, że działający jest winien skutków ubocznych swojego postępowania.

Hipoteza ta prowadziła do stwierdzenia, że zazwyczaj, kiedy ludzie uznają kogoś za winnego pewnej sytuacji, musieliby go też uznać za działającego intencjonalnie. Jednak w badaniach Nadelhoffera (por. przyp. 5) badani stwierdzili, że kierowca samochodu jest winny, mimo że nie działał intencjonalnie. Dodatkowo, jak zauważa Knobe (2006), według modelu zaproponowanego przez Mallego i Nelson przekonanie o tym, że działanie było intencjonalne, jest spowodowane orzekaniem winy. W związku z tym badani nie powinni byli go mieć, jeśli sytuacja dotyczyła działania, którego skutki nie są przedmiotem moralnej winy. Natomiast w przykła-

---

<sup>4</sup> Oceniając scenariusz: "Susan is the president of a major computer corporation. One day, her assistant comes to her and says, 'We are thinking of implementing a new program. If we actually do implement it, we will be increasing sales in Massachusetts but decreasing sales in New Jersey.' Susan thinks, 'According to my calculations, the losses we sustain in New Jersey should be a little bit smaller than the gains we make in Massachusetts. I guess the best course of action would be to approve the program.' 'All right', she says. 'Let's implement the program. So we'll be increasing sales in Massachusetts and decreasing sales in New Jersey.'" (Knobe, Mendlow, 2004), większość badanych stwierdziła, że Susan obniżyła poziom sprzedaży w New Jersey intencjonalnie (75%).

<sup>5</sup> Oceniając scenariusz: "Bob got rip-roaring drunk at a party after work. When the party ended, he stumbled to his car and started driving home. He was very drunk at the time — so drunk that he eventually lost control of his car, swerved into oncoming traffic, and killed a family of five" (Nadelhoffer 2004), większość badanych stwierdziła, że Bob nie doprowadził do śmierci rodziny intencjonalnie (92%).

dzie Knobego i Mendlowa (por. przyp. 4) badani ocenili, że działanie bohaterki, które miało stosunkowo niekorzystne skutki, było intencjonalne, mimo że było ono neutralne moralnie, i jak pokazały dalsze badania (Knobe 2004, Phelan, Sarkissian 2008), nie wzbudzało przekonania o winie. Podsumowując, teoria błędu jest niewystarczająca, ponieważ nie uwzględnia zjawisk opisanych w punktach III i IV, to znaczy tego, że badani mogą przypisać winę bez stwierdzenia, iż działanie było intencjonalne, oraz mogą ocenić działanie jako intencjonalne bez przypisania winy.

#### 4.2. Teoria decyzji

Inną teorię mającą tłumaczyć efekt Knobego zaproponował Machery (2008). Stwierdził, że dwie wykorzystane przez Knobego historyjki zasadniczo różnią się nie tylko aspektem moralnym działania prezesa, lecz także rozumowaniem, które musiał przeprowadzić prezes, aby podjąć decyzję. W przypadku scenariusza negatywnego badani postrzegają działanie prezesa jako rodzaj świadomej kalkulacji. Prezes zgadza się na koszty (pogorszenie stanu środowiska) w zamian za pożądany rezultat (większe zyski). Badani postrzegają kalkulację kosztów przeprowadzaną przez prezesa decydującego się na zaskodzenie środowisku w celu zwiększenia zysków jako działanie świadome. Zatem oceniają je jako intencjonalne. W historyjce dotyczącej poprawy stanu środowiska o takich kosztach nie ma mowy. Prezes otrzymuje pozytywną informację o spodziewanych zyskach i towarzyszącej temu poprawie stanu środowiska. Nie wymaga ona od niego kalkulacji, nie jest to informacja, która miałaby wpłynąć na zmianę decyzji i jako taka nie jest mu przedstawiana. Biorąc to wszystko pod uwagę, badani mogli postrzegać negatywne działanie prezesa jako bardziej świadome, bo wykalkulowane.

Machery twierdzi, że efekt Knobego nie ma nic wspólnego z przekonaniem o moralnej winie i wystąpiłby również w sytuacjach nienacechowanych moralnie. Byłoby to zgodne również z przykładem Knobego i Mendlowa, w którym mamy co prawda do czynienia z działaniem neutralnym moralnie, lecz jest ono wykalkulowane. Propozycji Machery'ego nie potwierdziły jednak kolejne eksperymenty. Ron Mallon (2008) przedstawił badanym dwie historyjki, w których wszystkie działania bohatera (zarówno te pozytywne, jak i negatywne moralnie) zostają przedstawione jako wymagające kalkulacji kosztów, a mimo to pojawia się asymetria w ocenie intencjonalności w zależności od tego, czy działanie było postrzegane przez czytelników jako dobre, czy jako złe. Podsumowując, asymetria w ocenie intencjonalności pojawia się w wynikach badań niezależnie od tego, czy decyzje podejmowane w scenariuszach są przedstawiane jako wymagające świadomej kalkulacji. Jest ona natomiast zależna od tego, czy efekty uboczne działań są przedstawiane jako pozytywne, czy negatywne.

## 5. TEORIA ODPOWIEDZIALNOŚCI A EFEKT KNOBEGO

Dwie przedstawione propozycje teoretyczne nie spełniają pierwszego warunku. Nie są dostatecznie ogólnymi teoriami asymetrii, ponieważ nie tłumaczą ogółu dostępnych wyników badań. W literaturze przedmiotu znaleźć można jednak projekt teoretyczny (Wright, Bengson 2009), który w przeciwieństwie do omówionych rozwiązań aspiruje do całościowego ujęcia problemu asymetrii i jest w stanie wyjaśnić wszystkie z zebranych dotychczas danych. Bazuje on na wprowadzeniu do rozważań pojęcia odpowiedzialności, które rozumiane jest szeroko, w szczególności nie ogranicza się do kategorii moralnych. Dzięki temu teoria ta może objąć przypadki asymetrii w stosunku do działań nienacechowanych moralnie ani emocjonalnie (por. punkt III).

Autorzy proponują teorię odpowiedzialności, zgodnie z którą asymetria powstaje na skutek powiązania dwóch pojęć: odpowiedzialności i intencjonalnego działania. Orzekanie o tym, czy działanie było intencjonalne, nie zależy jednak od orzekania winy (jak było w przypadku klasycznej teorii błędu), ale od orzekania odpowiedzialności.

Główna teza Wright i Bengsona głosi, że zaobserwowana empirycznie asymetria we wnioskowaniu na temat intencjonalności działań jest spowodowana asymetrią w przypisywaniu odpowiedzialności za te działania. Według autorów ludzie posiadają nierówne standardy przypisywania odpowiedzialności za złe lub naganne czyny (tzw. negatywna odpowiedzialność) i za czyny dobre lub godne pochwały (tzw. pozytywna odpowiedzialność). Zgodnie z tymi standardami przypisywanie negatywnej odpowiedzialności za czyn jest łatwe i częste, wymaga jedynie wiedzy, czy sprawca był świadomy możliwych skutków swojego zachowania. Przypisanie pozytywnej odpowiedzialności wymaga natomiast wiedzy, że działanie zostało podjęte ze względu na pewien dobry powód; wymaga wiedzy, że pozytywne działanie, dobre w skutkach, zostało podjęte w sposób zamierzony, a nie było dziełem przypadku. Dopiero w takiej sytuacji sprawca jest odpowiedzialny za nie i za jego pozytywnie skutki.

Wright i Bengson (2009: 27) twierdzą, że nie w każdym przypadku fakt przypisania odpowiedzialności sprawcy działania pociągnie za sobą ocenę, zgodnie z którą działanie to było intencjonalne. Sprawca może być odpowiedzialny również za działania nieintencjonalne, które zostały spowodowane przez niedopatrzanie (tak jest np. w wypadku pijanego kierowcy odpowiedzialnego za wypadek samochodowy), a także przez ignorancję czy niewiedzę. Innymi słowy teoria odpowiedzialności dopuszcza jednoczesne przypisanie komuś odpowiedzialności i stwierdzenie, że jego działanie było nieintencjonalne. Stwierdzenie, że działanie było intencjonalne, jest przy tym mniej prawdopodobne w razie nieprzypisania odpowiedzialności za czyn. Analiza korelacji, którą przeprowadzili autorzy, opierając się na własnych badaniach, wykazała, że ludzie byli bardziej skłonni uznawać czyn za intencjonalny, gdy jedno-



częściej oceniali go jako godny pochwały lub nagany, a więc gdy uważali, że jego sprawca ponosi odpowiedzialność<sup>6</sup>.

Zgodnie z teorią odpowiedzialności efekt Knobego powstaje, ponieważ badani dużo częściej obarczają sprawcę odpowiedzialnością w przypadku negatywnej historyjki niż w przypadku pozytywnej. W związku z tym są bardziej skory do przypisywania intencjonalności działaniom o niekorzystnych skutkach ubocznych. W ten sposób powstaje asymetria w ocenie:

1. W wypadku scenariusza ze szkodzeniem środowisku badani chętniej przypisują prezesowi zarządu negatywną odpowiedzialność, ponieważ jego wiedza o skutkach ubocznych (w tym wypadku o niszczeniu środowiska) jest wystarczającym warunkiem przypisania mu negatywnej odpowiedzialności, a co za tym idzie — uznania jego działania za intencjonalne.

2. W scenariuszu o polepszeniu stanu środowiska nie podaje się wystarczających powodów do przypisania pozytywnej odpowiedzialności, a więc do uznania działania za intencjonalne. Pomoc środowisku była konsekwencją, za którą prezes zarządu nie jest pozytywnie odpowiedzialny, ponieważ z historyjki nie wynika, że był to jeden z motywów jego działania (choć miał o skutkach działania pełną wiedzę).

Co więcej, zapytano badanych, *dlaczego* stwierdzili, że prezes zarządu intencjonalnie zaszkodził lub nieintencjonalnie pomógł środowisku (por. Beebe, Buckwalter 2010). Okazało się, że w wypadku stwierdzenia, iż działanie było intencjonalne, badani odwoływali się do wiedzy prezesa na temat konsekwencji jego działań. Natomiast opinię, że działanie nie było intencjonalne, uzasadniali tym, że pomoc środowisku nie była motywem postępowania prezesa.

Podsumowując, zwolennicy TO twierdzą, że orzekanie intencjonalności działania jest konsekwencją przypisania odpowiedzialności. Kiedy przypisujemy komuś odpowiedzialność, stwierdzamy też, że jego działanie było intencjonalne. Skoro częściej przypisujemy negatywną odpowiedzialność, to również częściej w negatywnie ocenianych sytuacjach będziemy rozstrzygać, że działanie było intencjonalne. Skutkiem może być występowanie efektu Knobego.

## 5.1. Teoria odpowiedzialności a inne zjawiska asymetrii

### 5.1.1. ASYMETRIA MIĘDZY ZDARZENIAMI NEUTRALNYMI A SILNIE NACECHOWANYMI MORALNIE

Zgodnie z punktem III ludzie mają tendencję do stwierdzania, że działania przypadkowe lub losowe są intencjonalne, o ile występują w kontekście silnie nacechowanym moralnie. Co ciekawe, asymetria występuje tutaj między losowymi, *neutral-*

---

<sup>6</sup> Autorzy nie formułują koniecznych warunków stwierdzenia, że dane działanie zostało wykonane intencjonalnie, rekonstruują jedynie pewne schematy rozumowań (Wright, Bengson 2009: 38).

*ny* moralnie zdarzeniami uznawanymi za nieintencjonalne (np. intencja wygrania zawodów oraz strzelenie przez przypadek w środek tarczy i tym samym wygranie zawodów; por. przyp. 3, sytuacja A) a losowymi zdarzeniami *nacechowanymi* moralnie — zarówno pozytywnie, jak i negatywnie — i uznawanymi za intencjonalne (por. przyp. 3, sytuacja B, zła moralnie: intencja zabicia ciotki oraz trafienie przez przypadek w ciotkę i tym samym zabicie jej; lub sytuacja C, oceniona jako pozytywna moralnie: intencja powstrzymania złego wroga, trafienie przez przypadek we wroga i zabicie go). Nie ma w tym wypadku asymetrii między zdarzeniami moralnie dobrymi a moralnie złymi. TO tłumaczy ten fakt, odwołując się do spostrzeżenia, że w wypadku wygrania gry losowej nie będziemy mówić o odpowiedzialności, a tym samym nie będziemy skłonni ocenić działania jako intencjonalnego.

### 5.1.2. PRZYPISANIE WINY A OCENA, CZY DZIAŁANIE BYŁO INTENCJONALNE

Według warunku zawartego w punkcie IV mogą zdarzyć się sytuacje, w których ludzie ocenią czyjeś działanie jako nieintencjonalne niezależnie od przypisania podmiotowi odpowiedzialności. Przykład z pijanym kierowcą, który w odczuciu badanych nieintencjonalnie spowodował wypadek i zabił innych ludzi, lecz mimo tego został uznany przez nich za winnego, może być problematyczny dla teorii błędu (Malle, Nelson 2003). Według niej przypisanie winy moralnej zwykle pociąga za sobą stwierdzenie, że działanie było intencjonalne. Natomiast zgodnie z TO można przypisać sprawcy odpowiedzialność za działanie, niekoniecznie stwierdzając, że było ono intencjonalne (mogło być również skutkiem ignorancji, niedopatrzania, niewiedzy itp.). Dlatego deklaracje badanych, że pijany kierowca nie przyczynił się do śmierci rodziny intencjonalnie, a jednocześnie ponosi za nią odpowiedzialność, nie są kontrprzykładem dla teorii odpowiedzialności.

### 5.1.3. ASYMETRIA W PRZYPADKACH NIENACECHOWANYCH MORALNIE

Więcej uwagi należy poświęcić wynikom eksperymentu opisanego w przypisie do punktu III (przyp. 4). Są one najbardziej problematyczne dla wszystkich teorii wyjaśniających asymetrię, w tym również dla TO. Opisana asymetria występuje między oceną działań, które w ogóle nie są nacechowane moralnie, jak na przykład zwiększanie lub zmniejszanie sprzedaży, a zatem są działaniami, które można oceniać jedynie jako relatywnie korzystne lub niekorzystne. Jeśli Knobe twierdzi, że nasze sądy moralne wpływają na nasze sądy o intencjonalności działań, to omawiany przypadek wykracza poza jego tezę. Jest to bowiem sytuacja, w której nie mamy do czynienia z rozumowaniem moralnym. Knobe nie zbadał ponadto, czy efekt asymetrii wystąpiłby również przy nieco odmienniejszej historii: nie sprawdził, czy doprowadzenie do wzrostu sprzedaży będzie chętniej uznawane za nieintencjonalne. Udowodniły to dopiero badania Wright i Bengsona (2009: 39). Wyniki pokazują, że w takim wypadku badani oceniają wzrost sprzedaży jako znacząco mniej intencjonalny.

Tłumacząc wyniki tych badań przez odwołanie do TO, należy zauważyć, że chociaż menadżer nie jest winny w sensie moralnym, to ponosi odpowiedzialność za

swoje działanie. W wystarczającym stopniu tłumaczy to ocenę jego działania jako intencjonalnego.

## 5.2. Krytyczna analiza teorii odpowiedzialności

Mimo że teoria odpowiedzialności spełnia warunek ogólności, który nałożyliśmy na teorię dobrze tłumaczącą asymetrię w ocenie, czy działanie było intencjonalne, nie jest ona dostatecznie dobrze ugruntowana empirycznie. Według jej autorów wnioskowanie o intencjonalności jest następstwem przypisania odpowiedzialności. To oznacza, że stwierdzenie, iż działanie było intencjonalne, powinno być pociągane (z reguły) nie tylko przez przypisanie odpowiedzialności negatywnej, lecz także *pozytywnej*. Badania dotyczące efektu Knobego są interpretowane jako potwierdzenie TO, ponieważ zakłada się, że uznanie, iż przewodniczący zarządu był odpowiedzialny za uszkodzenie środowisku, wpłynęło na rozstrzygnięcie, że było to działanie intencjonalne. W literaturze przedmiotu przeważnie przedstawiane są badania opisujące tylko związek między przypisaniem negatywnej odpowiedzialności a stwierdzeniem, że działanie było intencjonalne. Niewiele jest badań, które pokazywałyby, że przekonanie o intencjonalności może być następstwem przypisania odpowiedzialności pozytywnej.

Zgodnie z TO asymetria w ocenie, czy działanie było intencjonalne (czyli czy zaszedł efekt Knobego), nie powinna wystąpić, gdy badany ma podstawy, aby w jednym z dwóch analogicznych scenariuszy przypisać sprawcy odpowiedzialność negatywną, a w drugim — pozytywną. Powinien odpowiedzieć, że działanie było intencjonalne w obu wypadkach, i pozytywnym, i negatywnym. Jak się wydaje, wyniki badania Wright i Bengsona (2009: 39) mogą stanowić kontrprzykład dla takich przewidywań. Autorzy użyli następujących historyjek:

GRUPA 1. SPADEK: Wiceprezes spółki poszedł do przewodniczącego zarządu i powiedział: „Myślimy o rozpoczęciu nowego projektu. To pomoże nam zwiększyć zyski, ale także zmniejszy sprzedaż w New Jersey”. Przewodniczący odparł: „W ogóle mnie nie obchodzi zmniejszenie sprzedaży w New Jersey. Chcę tylko osiągnąć jak największe zyski. Zaczniemy nowy projekt”. Zaczęli nowy projekt. Zyski rzeczywiście wzrosły, a sprzedaż w New Jersey spadła.

GRUPA 2. WZROST: Wiceprezes spółki poszedł do przewodniczącego zarządu i powiedział: „Myślimy o rozpoczęciu nowego projektu. To pomoże nam zwiększyć zyski, ale także zwiększy sprzedaż w New Jersey”. Przewodniczący odparł: „W ogóle mnie nie obchodzi zwiększenie sprzedaży w New Jersey. Chcę tylko osiągnąć jak największe zyski. Zaczniemy nowy projekt”. Zaczęli nowy projekt. I rzeczywiście, zyski wzrosły, tak jak i sprzedaż w New Jersey.

Analiza wyników pokazała, że badani częściej twierdzili, iż prezes zarządu intencjonalnie doprowadził do spadku sprzedaży w New Jersey, niż intencjonalnie doprowadził do wzrostu sprzedaży w New Jersey. Wydaje się jednak, że wystąpienie efektu

Knobego w opisywanej historyjce jest niezgodne z TO. Według TO asymetria nie powinna zachodzić, ponieważ zarówno w scenariuszu dotyczącym wzrostu, jak i spadku sprzedaży mamy do czynienia z działaniem na podstawie dobrego powodu i świadomej kalkulacji. TO przewiduje przypisanie negatywnej odpowiedzialności za spadek sprzedaży, ale bynajmniej nie wyklucza przypisania pozytywnej odpowiedzialności za wzrost sprzedaży. Tym samym nie tłumaczy, dlaczego występuje asymetria.

W scenariuszu ze wzrostem sprzedaży (Grupa 2) prezes zarządu mówi wprawdzie: „W ogóle mnie nie obchodzi zwiększenie sprzedaży w New Jersey. Chcę tylko osiągnąć jak największe zyski”, co mogłoby wskazywać, że tak jak we wcześniejszych badaniach, również tu nie są spełnione kryteria przypisania *pozytywnej* odpowiedzialności, ponieważ nie ma dobrego motywu działania. Należałoby jednak wziąć pod uwagę, że wzrost sprzedaży w New Jersey skądinąd przyczyni się do wzrostu ogólnych zysków. Głównym motywem działań prezesa jest chęć zwiększenia ogólnych zysków i jest to jednocześnie powód, dla którego powinien chcieć zwiększyć zyski w New Jersey. Innymi słowy, wzrost sprzedaży był efektem ubocznym, który był zgodny z celem głównym. Jeśli prezes podjął działania zmierzające do celu głównego (tu: zwiększenia zysków) z dobrego powodu, to ten sam powód powinien być motywem zwiększenia sprzedaży w New Jersey (czyli efektu ubocznego). Zatem oczekiwalibyśmy przypisania prezesowi pozytywnej odpowiedzialności za wzrost sprzedaży w New Jersey, ponieważ miał on dobry powód do działania. Można oczywiście polemizować z taką interpretacją badania. Być może struktura historyjki jest na tyle zawila, że badani raczej koncentrują się na słowach „w ogóle mnie nie obchodzi”, a nie zastanawiają się, czy zwiększenie sprzedaży w New Jersey jednocześnie zwiększy zyski firmy.

Trudność, która wiąże się ze zbadaniem związku między pozytywną odpowiedzialnością a intencjonalnością działań, polega na tym, że nie wiadomo dokładnie, jak sprecyzować kryteria przypisywania „pozytywnej odpowiedzialności” i kiedy faktycznie działający ma „dobry powód” do działania. Zarzut wobec teorii odpowiedzialności dotyczyłby zatem nieprecyzyjności używanej terminologii, co przyczynia się do powstania problemu operacjonalizacji teorii w konkretnych badaniach.

Dobrym przykładem, który może być interpretowany jako potwierdzenie TO w omawianym zakresie, mogą być badania nad „problemem Butlera” (por. przyp. 3). W momencie, w którym badani mieli możliwość przypisania zarówno pozytywnej odpowiedzialności (uratowanie życia ludzi), jak i negatywnej (odebranie życia drugiej osobie), różnica w asymetrii nie wystąpiła, widoczna była natomiast przy ocenie historyjki analogicznej, lecz moralnie neutralnej. Z uwagi na wyniki potwierdzające przewidywania autorów TO, jak i takie, które zdają się im przeczyć (opisane wyżej badanie efektu Knobego przez Wright i Bengsona), warto przyjąć określone i jasne kryteria przypisywania zarówno pozytywnej, jak i negatywnej odpowiedzialności. Należałoby również zaproponować kolejne badania, które zweryfikowałyby propozycje autorów TO, szczególnie w odniesieniu do pozytywnej odpowiedzialności, i wyjaśniły niezgodności w istniejących obecnie wynikach.

### 5.3. Wyzwania teorii odpowiedzialności

Przedstawiliśmy i przeprowadziliśmy krytyczną analizę teorii odpowiedzialności jako całościowej propozycji wyjaśniającej asymetrię w orzekaniu intencjonalności działań, zaobserwowaną w różnych badaniach empirycznych z zakresu filozofii eksperymentalnej.

Najbardziej znanym przykładem asymetrii jest efekt Knobego, jednak w kompleksowym podejściu do problemu asymetrii należy uwzględnić również inne wyniki badań empirycznych na ten temat. Zestawienie najważniejszych opisanych w literaturze zjawisk asymetrii pozwoliło pokazać, że wybrane propozycje teoretycznie nie są w stanie spójnie odnieść się do wszystkich dostępnych danych. Starałam się wyjaśnić, w jaki sposób teoria odpowiedzialności może spełnić takie zadanie. Autorzy TO twierdzą, że:

(i) Asymetria w orzekaniu intencjonalności działań powstaje ze względu na asymetrię w przypisywaniu odpowiedzialności.

(ii) Asymetria w przypisywaniu odpowiedzialności jest wynikiem stosowania nierównych kryteriów wobec zdarzeń pozytywnych i negatywnych.

Przewaga TO nad innymi teoriami polega więc na wyjaśnianiu relatywnie dużej liczby różnych zjawisk związanych z asymetrią, nie tylko tych, które są łączone ze zdarzeniami moralnie nacechowanymi. Główną wadą TO jest niedostatecznie potwierdzona empirycznie skuteczność w przewidywaniu decyzji badanych. W szczególności, poza eksperymentami dotyczącymi „problemu Butlera” niewiele jest przekonujących badań, które stwierdzałyby brak asymetrii w wypadku takich modyfikacji historyjek, które umożliwiają badanym przypisanie *pozytywnej* odpowiedzialności. Zaprojektowanie takiego badania może wiązać się z problemami terminologicznymi i pojęciowymi, np. ze sprecyzowaniem, kiedy mamy do czynienia z pozytywną odpowiedzialnością.

### BIBLIOGRAFIA

- Adams F., Steadman A. (2004), *Intentional Action in Ordinary Language. Core Concept or Pragmatic Understanding?*, „Analysis” 64(282), 173-181.
- Beebe J., Buckwalter W. (2010), *The Epistemic Side-Effect Effect*, „Mind & Language” 25(4), 474-498.
- Butler R. (1978), *Report on Analysis ‘Problem’ no. 16*, „Analysis” 38(3), 113-14.
- Knobe J. (2003), *Intentional Action and Side Effects in Ordinary Language*, „Analysis” 63(3), 190-194.
- Knobe J. (2004), *What is Experimental Philosophy?*, „The Philosophers’ Magazine” 28.
- Knobe J. (2006), *The Concept of Intentional Action. A Case Study in the Uses of Folk Psychology*, „Philosophical Studies” 130(2), 203-231.
- Knobe J. (2007), *Reason Explanation in Folk Psychology*, „Midwest Studies in Philosophy” 31(1), 90-106.

- Knobe J., Burra A. (2006), *The Folk Concepts of Intention and Intentional Action. A Cross-Cultural Study*, „Journal of Culture and Cognition” 6(1-2), 113-132.
- Knobe J., Mendlow G. (2004), *The Good, the Bad, and the Blameworthy. Understanding the Role of Evaluative Reasoning in Folk Psychology*, „Journal of Theoretical and Philosophical Psychology” 24(2), 252-258.
- Knobe J., Nichols S. (2008), *An Experimental Philosophy Manifesto* [w:] *Experimental Philosophy*, J. Knobe, S. Nichols (red.), Oxford: Oxford University Press, 3-14.
- Machery E. (2008), *The Folk Concept of Intentional Action. Philosophical and Experimental Issues*, „Mind & Language” 23(2), 165-189.
- Malle B. F., Nelson S. E. (2003), *Judging Mens Rea. The Tension Between Folk Concepts and Legal Concepts of Intentionality*, „Behavioral Sciences & the Law” 21(5), 563-580.
- Mallon R. (2008), *Knobe versus Machery. Testing the Trade-Off Hypothesis*, „Mind & Language” 23(2), 247-255.
- Nadelhoffer T. (2004), *The Butler Problem Revisited*, „Analysis” 64(3), 277-284.
- Phelan M., Sarkissian H. (2008), *The Folk Strike Back. Or, Why You Didn't Do It Intentionally, Though It Was Bad and You Knew It*, „Philosophical Studies” 138(2), 291-298.
- Wright J. C., Bengson J. (2009), *Asymmetries in Judgments of Responsibility and Intentional Action*, „Mind & Language”, 24(1): 24-50.