

WIKTOR PIOTROWSKI*

PRZYPISYWANIE POSTAW W EFEKCIE KNOBE'A**

Abstract

ATTITUDE ASCRIPTION IN THE KNOBE EFFECT

Joshua Knobe (2003a) showed that people are much more inclined to attribute intentionality to an agent when a side effect of her action is harmful than when it is beneficial. This asymmetry came to be known as the Knobe Effect. A number of theories tried to explain this phenomenon, including the so-called "moral hypothesis". This article presents a survey, based on the metaethical concept of attitude, that aims to make the hypothesis more precise. It is designed to test attitude ascription in the original Knobe's scenarios and compare it with the asymmetry in intentionality attribution. After presenting the survey and its results, I concentrate on various interpretations of the data and sketch a possible direction for further examination of the moral hypothesis.

Keywords: attitude, approval, disapproval, experimental philosophy, Knobe Effect, moral hypothesis

Jak pokazuje tzw. efekt Knobe'a, jesteśmy skłonni do przypisywania intencjonalności działaniom zależnie od ich statusu normatywnego. „Hipoteza moralna” to jedna z prób wyjaśnienia tego zjawiska. Według jej zwolenników skłonność do określania jakiegoś działania jako intencjonalnego częściowo wynika z chęci oceny jego wymiaru moralnego (Harman 1976, Knobe 2003a, 2006, Nadelhoffer 2004): respondenci oceniają działanie jako intencjonalne, ponieważ chcą dokonać pewnego rodzaju oceny moralnej, lecz konstrukcja oryginalnego badania Knobe'a nie dostarcza im takiej możliwości. Przypisywanie intencjonalności jest jednym z elementów oceny moralnej, a w szcze-

* Instytut Filozofii, Uniwersytet Warszawski, ul. Krakowskie Przedmieście 3, 00-927 Warszawa, wik.piotrowski@gmail.com.

** Artykuł powstał podczas prac seminarium „Studium teoretyczne i eksperymentalne efektu Knobe'a i problemu Butlera”, będącego częścią projektu finansowanego ze środków Narodowego Centrum Nauki (2013/11/B/HS1/03939).

gólności nieintencjonalność działania często osłabia lub niweluje negatywną ocenę danego czynu. Dlatego też respondenci z braku środków do dokonania pełnej oceny moralnej używają normatywnie nacechowanej koncepcji intencjonalności, zastępującej niejako ocenę moralną. Badani są świadomi, że intencjonalność pociąga za sobą odpowiedzialność rozumianą w sposób ściśle moralny. Intencjonalność jest w tym wypadku rozumiana jako jeden z terminów „psychologii potocznej” (*folk psychology*) i jest istotnie zależna od oceny moralnej (Knobe, Mendlow 2004).

Celem artykułu jest doprecyzowanie tak rozumianej hipotezy moralnej. Termin „dokonanie oceny moralnej” jest bowiem zbyt ogólny i nieprecyzyjny, a przez to pozbawiony wartości eksplanacyjnej. Pytanie o to, czy dane działanie jest dobre lub złe, powinno zostać zastąpione pytaniami o ocenę konsekwencji działań, o postawę podmiotów działania, o ich charakter lub o inny bardziej szczegółowy element oceny moralnej. W części 2 przedstawiam badania sprawdzające, w jaki sposób respondenci oceniają postawę opisaną w scenariuszu Knobe’a oraz jak należy ją ocenić z perspektywy moralnej. Moim celem nie jest zaproponowanie kompletnego wyjaśnienia efektu Knobe’a, lecz jedynie wyznaczenie kierunku badań oraz postawienie pierwszego kroku.

1. HIPOTEZA MORALNA – WARIANT „NAGANNY I GODNY POCHWAŁY”¹

Joshua Knobe (2003a) opisał następujące badanie: 78 osobom spacerującym po parku na Manhattanie została przedstawiona jedna historyjka w dwóch wersjach, nazywanych *harm condition* (warunek szkodzenia, S) i *help condition* (warunek pomagania, P).

WARUNEK SZKODZENIA	WARUNEK POMAGANIA
Wicedyrektor zwraca się do dyrektora pewnej firmy: „Myślmy o wdrożeniu nowego programu. Pozwoli nam zwiększyć zyski, ale zaszkodzi środowisku”.	Wicedyrektor zwraca się do dyrektora pewnej firmy: „Myślmy o wdrożeniu nowego programu. Pozwoli nam zwiększyć zyski i pomoże środowisku”.
Dyrektor odpowiada: „Nie obchodzi mnie szkodzenie środowisku. Chcę tylko zwiększyć zyski. Wdrażamy program”.	Dyrektor odpowiada: „Nie obchodzi mnie pomaganie środowisku. Chcę tylko zwiększyć zyski. Wdrażamy program”.

¹ Ang. *blameworthy* i *praiseworthy*. W dalszej części tekstu terminy te zostają oddane jako „przypisanie winy” oraz „przypisanie zasługi”, ponieważ takie sformułowanie najlepiej oddaje intencje autorów. Uściślenie tej terminologii zawdzięczam recenzentowi artykułu.

Program został wdrożony i rzeczywiście zaszkodził środowisku.	Program został wdrożony i rzeczywiście pomógł środowisku.
Czy dyrektor intencjonalnie (<i>intentionally</i>) zaszkodził środowisku?	Czy dyrektor intencjonalnie pomógł środowisku?
TAK – 82%	TAK – 23%
NIE – 18%	NIE – 77%

Tabela 1. Polskie tłumaczenie scenariuszy wykorzystanych w (Knobe 2003a) oraz uzyskane przez Knobe'a wyniki

Wyniki badania pokazały, że gdy skutek uboczny działania dyrektora był negatywny, większość respondentów przypisywała mu intencjonalność, a zupełnie inaczej było w warunku pomagania. Wynik ten jest o tyle zaskakujący, o ile standardowe koncepcje intencjonalności (por. Davidson 1963, Anscombe 1963) zakładają, że intencjonalność działania jest związana wyłącznie z pragnieniami i przekonaniem działającego podmiotu. W koncepcji Davidsona polega ona na działaniu ze względu na pewną rację (*reason*), która składa się z połączenia pragnienia i przekonania. Znajomość nadrzędnej racji (*primary reason*) działania jest tożsama ze znajomością intencji, z jaką zostało przeprowadzone (Davidson 1963: 688-689). Dlatego gdyby respondenci oceniali przedstawioną przez Knobe'a sytuację zgodnie z koncepcją Davidsona, byłiby zobowiązani twierdzić, że to nieprawda, jakoby dyrektor intencjonalnie szkodził środowisku. Nie miał bowiem odpowiednich pragnień i przekonań. Podobnie jest z koncepcją Anscombe, która charakteryzuje działanie intencjonalne jako działanie, w stosunku do którego sensownie można postawić pytanie „Dlaczego?” (Anscombe 1963: 24). Według Anscombe (podobnie jak u Davidsona), aby uznać działanie za intencjonalne, odpowiedź na pytanie „Dlaczego?” powinna odwoływać się do pewnych racji związanych z przekonaniem i pragnieniami podmiotu.

Mimo licznych różnic koncepcje te mają wspólny rys: traktują działanie intencjonalne jako działanie wymagające wyjaśnienia, podania racji, na którą składają się pragnienia i przekonania podmiotu. Z tego powodu są zobowiązane traktować scenariusze Knobe'a jako przykłady działania nieintencjonalnego: dyrektor nie miał racji po temu, by szkodzić czy też pomagać środowisku, nie może też odpowiedzieć na pytanie „Dlaczego?”. Standardowe ujęcia intencjonalności nie są więc w stanie wyjaśnić efektu Knobe'a: nie potrafią wytłumaczyć, skąd bierze się tak duża rozbieżność w przypisywaniu intencjonalności w obydwu scenariuszach, skoro różnią się one jedynie moralną oceną skutku ubocznego.

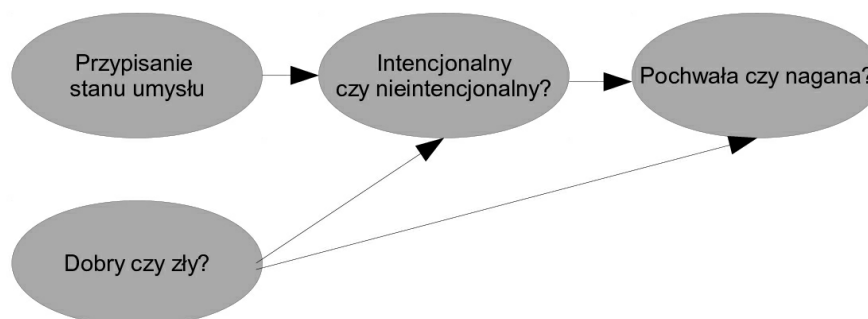
Badanie Knobe'a dostarczyło filozofii danych, których wyjaśnienie stanowi problem dla większości współczesnych teorii działania intencjonalnego.

Powstały liczne hipotezy, które miały za zadanie wyjaśnić widoczne w eksperymencie Knobe'a asymetryczne przypisywanie intencjonalności: teoria błędu (Malle, Nelson 2003), teoria decyzji (Machery 2008), teoria odpowiedzialności (Wright, Bengson 2009), hipoteza pragmatyczna (Adams, Steadman 2004), teoria działania intencjonalnego Hindriksa (2008) oraz hipoteza zaniechaniowa (Paprzycka 2014) i inne². W artykule omówię hipotezę moralną, zaproponowaną przez samego Knobe'a (2003a, b), która jest prawdopodobnie najszerzej komentowaną próbą wytłumaczenia tego zjawiska.

Początkowo Knobe (2003a) zestawiał różnicę w przypisywaniu intencjonalności z różnicą co do ewentualnego przypisania winy lub zasługi dyrektorowi. Okazywało się bowiem, że respondenci często przypisywali dyrektorowi winę w przypadku S ($M = 48$), natomiast znacznie rzadziej przypisywali dyrektorowi zasługę w przypadku P ($M = 14$). Hipotezą Knobe'a było więc wówczas założenie, że różnica w przypisywaniu intencjonalności wynika z tego, że ludzie niechętnie przypisują zasługę dyrektorowi, gdy efekt uboczny jego działania jest oceniany jako dobry, za to chętnie przypisują mu winę, gdy efekt jego działania jest oceniany jako zły. Dlatego też różnica w przypisywaniu intencjonalności związana miała być z różnicą w przypisywaniu winy lub zasługi.

Hipoteza ta została zarzucona w artykule z 2004 r. (Knobe, Mendlow 2004) w wyniku polemiki z Nadelhofferem (2004). Knobe i Mendlow wysunęli przypuszczenie, że ocena intencjonalności zależy od oceny moralnej, nie jest jednak zapośredniczona przez przypisywanie winy lub zasługi: najpierw przypisany zostaje pewien stan umysłu (np. przekonania lub pragnienia), następnie przypisuje się intencjonalność — na co wpływa ocena moralna — a dopiero na końcu następuje przypisanie winy lub zasługi, które również jest silnie zależne od oceny moralnej. Głównym założeniem tej konstrukcji było uznanie intencjonalności za funkcjonalne pojęcie psychologii potocznej. Autorzy twierdzą, że intencjonalność jest użytecznym pojęciem, które służy do tego, by zająć odpowiednią postawę względem całej sytuacji. Przypisywanie intencjonalności jest determinowane przez to, czy w danej sytuacji bardziej odpowiednie jest przypisanie winy, czy zasługi. Knobe i Mendlow twierdzą wbrew Nadelhofferowi, że przypisanie intencjonalności odbywa się przed przypisaniem winy lub zasługi, lecz pozostaje jednocześnie pod wpływem oceny moralnej. Całość tego procesu autorzy przedstawili za pomocą następującego schematu:

² Pełniejszy i bardziej szczegółowy przegląd alternatywnych wyjaśnień efektu Knobe'a można znaleźć w artykułach Knobe'a (2010) i Dębskiej (2013).



Rysunek 1. Rola intencjonalności w psychologii potocznej (Knobe, Mendlow 2004)

Zgodnie z hipotezą moralną intencjonalność służy nie tyle przewidywaniu i wyjaśnianiu zachowania, ile jego ocenianiu (Knobe 2003b: 321). Od moralnej oceny sytuacji będzie zatem zależało zastosowanie pojęcia intencjonalności. Jeżeli respondenci uważają, że dane postępowanie jest moralnie złe, to uznają je za intencjonalne, ponieważ stanowi to jedyny dostępny im w badaniu środek wyrażenia dezaprobaty. Hipoteza Knobe'a i Mendlowa brzmiała więc: zastosowanie pojęcia intencjonalności zależy od tego, czy daną sytuację uważa się za dobrą, czy złą.

Spróbuję pokazać, że jest to hipoteza tyleż właściwa, co niejasna. Pojęcie „dobra i zła danej sytuacji” niczego nie tłumaczy, a zaproponowany schemat, choć stanowi dobry punkt wyjścia, jest zbyt uproszczony. Przedstawię badania, które miały za zadanie rozbudowę i modyfikację schematu Knobe'a i Mendlowa oraz wskazanie kierunku, w którym mogłaby rozwijać się omawiana wersja hipotezy moralnej.

2. OCENA POSTAWY JAKO JEDEN Z ELEMENTÓW OCENY MORALNEJ

Pojęcie postawy (*attitude*) użyte w przedstawianym tu badaniu jest zaczerpnięte z tradycji metaetycznej. Zostało rozpowszechnione w teoriach emotywistycznych, głównie za sprawą Stevenson (1963), dla którego postawa to „wszelka dyspozycja psychiczna bycia za czymś lub przeciwko czemuś” (Stevenson 1975: 114). Choć Stevenson nie był autorem tego określenia, to w znacznym stopniu pod jego wpływem zastąpiło ono wcześniej używane — np. przez Perry'ego (1926) — określenie *interest*, będące angielskim odpo-

wiednikiem niemieckiego terminu *Interesse*. Pojęcia *attitude* i *interest*, gdy używane są w sensie technicznym, można uznać za synonimy: przeciwstawia się je pojęciu przekonania lub wiedzy. Odnoszą się do niekognitywnego stanu umysłu, który jest jednocześnie blisko związany z siłą motywacyjną. Jak pisze Satris (1987), oznaczają szeroko rozumianą dyspozycją behawioralną — nie tylko do działania, lecz także do przeżywania czegoś w określony sposób, do odczuwania takich, a nie innych emocji. Postawa jako niekognitywny stan umysłu koncentruje się wokół chcenia lub pragnienia czegoś. Bycie za czymś można zatem przełożyć jako pragnienie urzeczywistnienia czegoś, natomiast bycie przeciw czemuś jest rozumiane jako pragnienie nieurzeczywistnienia czegoś. Technicznymi terminami, których Stevenson używa do określenia tych dwóch rodzajów postaw, są „aprobata” i „dezaprobata”. W badaniu zostały one wykorzystane do opisanego stanu umysłu dyrektora oraz do zajęcia postawy wobec jego stanu umysłu: w pierwszym pytaniu badani mieli ocenić postawę dyrektora jako jego dyspozycję do działania, natomiast w drugim mieli ocenić postawę dyrektora w sposób niezwiązany bezpośrednio z żadnym określonym działaniem z jego strony.

Badanie zostało przeprowadzone w 2015 r. za pośrednictwem platformy LimeSurvey. Łącznie wzięło w nim udział 113 osób losowo wybranych spośród chętnych.

2.1. CEL BADANIA

Celem badania było rozstrzygnięcie, czego dotyczy ocena moralna w efekcie Knobe’a, a dokładniej sprawdzenie, czy dotyczy oceny postawy dyrektora. W badaniu przyjęto więc założenie hipotezy moralnej, zgodnie z którym respondenci przypisują działaniom dyrektora intencjonalność, ponieważ chcą dokonać oceny moralnej. Badanie miało doprecyzować hipotezę — w tym celu badani byli pytani o ocenę postawy dyrektora.

2.2. STRUKTURA BADANIA

Badani byli losowo przydzielani do jednej z dwóch grup, do warunku S lub P. W każdej z tych grup zapoznawali się z oryginalnymi scenariuszami Knobe’a, które pomijały jedynie opis rezultatów działania³, a następnie mieli udzielić

³ Głównym celem badania było skupienie uwagi badanych na samej postawie dyrektora. Fragment mówiący o rzeczywistych konsekwencjach tych działań sprawiał, że przypisanie postawy dyrektorowi, jak i jej ocena mogły zostać zakłócone przez wiedzę o skutkach działania. Aby wyizolować samo przypisanie stanu umysłu i jego ocenę, w badaniu usunięto więc zdanie: „Program został wdrożony. Zgodnie z oczekiwaniami zaszkodził/pomógł śro-

odpowiedzi na dwa (identyczne dla obu grup) pytania. Całość badania wyglądała następująco:

Wicedyrektor zwraca się do dyrektora pewnej firmy: „Myślimy o wdrożeniu nowego programu. Pozwoli nam zwiększyć zyski, ale [i] zaszkodzi [pomocze] środowisku”. Dyrektor odpowiada: „Nie obchodzi mnie szkodzenie [pomaganie] środowisku. Chcę tylko zwiększyć zyski. Wdrażamy program”.

Pytanie 1. Które wyrażenie najlepiej opisuje postawę dyrektora?

- A1. Dezaprobuje pomaganie środowisku.
- A2. Aprobuję pomaganie środowisku.
- A3. Dezaprobuje szkodzenie środowisku.
- A4. Aprobuję szkodzenie środowisku.
- A5. Jest obojętny wobec środowiska.

Pytanie 2. Jak oceniasz postawę dyrektora?

- A1. Dezaprobuję postawę dyrektora.
- A2. Aprobuję postawę dyrektora.
- A3. Jestem obojętny wobec postawy dyrektora.

2.3. HIPOTEZY BADAWCZE

Pytanie pierwsze miało na celu sprawdzenie, jaką postawę badani przypisują dyrektorowi, czyli jak traktują stwierdzenie dyrektora „nie obchodzi mnie pomaganie/szkodzenie środowisku”. Zgodnie z badaną hipotezą respondenci w większości powinni zaznaczać odpowiedź A5: „[Dyrektor] jest obojętny wobec środowiska”. Przyjmuję zatem, że najbardziej intuicyjnym ro-

dowisku”. James Beebe, podczas warsztatów organizowanych przez Instytut Filozofii UW w maju 2015 r., w dyskusji nad wynikami opisywanych w artykule badań, wysunął hipotezę, że brak tego zdania może znacząco wpływać na różnice w odpowiedziach. W celu zweryfikowania tej hipotezy przeprowadzone zostały badania, w których zdanie nie zostało usunięte. Badanie było prowadzone od października 2015 do marca 2016 za pośrednictwem platformy LimeSurvey. Łącznie wzięły w nim udział 53 osoby, losowo wybrane spośród chętnych. Hipoteza Beebe’ego głosiła, że ankietowani inaczej będą oceniać postawę dyrektora, jeżeli poda im się informację o rzeczywistych skutkach jego działania. Oznaczałoby to, że różnica między oczekiwanym a rzeczywistym skutkiem wpływa na ocenę postawy. Wyniki badania z zachowaną informacją o skutkach działania dyrektora powinny więc w stopniu statystycznie istotnym różnić się od wyników badania, w którym brak było tej informacji. Tak się jednak nie stało. Wyniki badania weryfikującego hipotezę Beebe’ego były niemalże identyczne z wynikami badań omawianych w tym artykule, dlatego też usunięcie zdania o rzeczywistych skutkach działania nie może być traktowane jako element zakłócający otrzymane wyniki.

zumieniem stwierdzenia „nie obchodzi mnie” jest potraktowanie go jako wyrazu obojętności. Możliwe jest też uznanie go za wyraz bardzo konkretnej postawy dezaprobującej. Wydaje się jednak, że takie odczytanie wymaga pewnych dodatkowych informacji dotyczących charakteru dyrektora, tonu głosu, z jakim wypowiedzane jest to stwierdzenie itp. Ponieważ respondenci otrzymują bardzo niewiele informacji o dyrektorze, zakładałem, że częściej traktują jego wypowiedź jako wyraz obojętności niż dezaprobaty.

Pytanie drugie miało sprawdzić, jak badani oceniają postawę dyrektora. Stawiałem hipotezę, że w wypadku S badani będą wyrażali dezaprobatę wobec postawy dyrektora, natomiast w P będą aprobowali jego postawę lub pozostaną wobec niej obojętni. Warunek S i warunek P różnią się bowiem jedynie oczekiwanym skutkiem ubocznym działania. Zgodnie z hipotezą moralną scenariusz Knobe'a skłania respondentów do dokonania oceny moralnej i w oryginalnym badaniu służy im do tego przypisanie intencjonalności. Przyjmuję zatem, że wyrażenie dezaprobaty lub aprobaty wobec czyjejs postawy również jest sposobem dokonywania pewnego rodzaju oceny moralnej. Dlatego też, jeżeli badani posługiwali się przypisaniem intencjonalności w celu dokonania oceny moralnej i w ocenie tej widać było wyraźną asymetrię, to w omawianym badaniu powinni posłużyć się wyrażeniem aprobaty i dezaprobaty, które również charakteryzować się będzie asymetrią. Innymi słowy, zgodnie z ogólnym założeniem asymetrii w przypisywaniu intencjonalności działania powinna przekładać się na asymetryczną ocenę postawy dyrektora. Hipoteza ta okazała się błędna.

2.4. WYNIKI BADANIA

Odpowiedzi na pierwsze pytanie dla S wyglądały następująco:

ODPOWIEDŹ	POSTAWA DYREKTORA	LICZBA ODPOWIEDZI
A1	Dezaprobuje pomaganie	3
A2	Dezaprobuje szkodzenie	1
A3	Aprobuje szkodzenie	14
A4	Aprobuje pomaganie	0
A5	Jest obojętny	46
Razem:		64

Tabela 2. Wyniki uzyskane dla pierwszego pytania w warunku szkodzenia

Niewielkie różnice widać w przypadku scenariusza P:

ODPOWIEDŹ	POSTAWA DYREKTORA	LICZBA ODPOWIEDZI
A1	Dezaprobuję pomaganie	7
A2	Dezaprobuję szkodzenie	0
A3	Aprobuję szkodzenie	5
A4	Aprobuję pomaganie	0
A5	Jest obojętny	41
Razem:		53

Tabela 3. Wyniki uzyskane dla pierwszego pytania w warunku pomagania

Odpowiedzi na drugie pytanie wyglądają następująco (osobno dla warunków S i P):

ODPOWIEDŹ	POSTAWA RESPONDENTA	LICZBA ODPOWIEDZI
A1	Aprobuję postawę dyrektora	2
A2	Dezaprobuję postawę dyrektora	50
A3	Jestem obojętny wobec postawy dyrektora	12
Razem:		64

Tabela 4. Wyniki uzyskane dla drugiego pytania w warunku szkodzenia

ODPOWIEDŹ	POSTAWA RESPONDENTA	LICZBA ODPOWIEDZI
A1	Aprobuję postawę dyrektora	2
A2	Dezaprobuję postawę dyrektora	39
A3	Jestem obojętny wobec postawy dyrektora	12
Razem:		53

Tabela 5. Wyniki uzyskane dla drugiego pytania w warunku pomagania

2.5. INTERPRETACJA WYNIKÓW

Zgodnie z pierwszą hipotezą badawczą zdecydowana większość ankietowanych traktuje wypowiedź dyrektora jako wyraz postawy obojętnej wobec środowiska. W przypadku pierwszego pytania największa różnica między scenariuszami zachodzi między odpowiedziami „aprobuje szkodzenie”. W scenariuszu S czternaście osób uznało, że wypowiedź dyrektora jest wyrazem pewnego rodzaju aprobaty, natomiast w P uznało tak tylko pięć osób. Jest to za-

skakujące z punktu widzenia tradycji metaetycznej, zgodnie z którą aprobata jest stanem umysłu wymagającym pewnego pozytywnego nakierowania na swój przedmiot, podczas gdy dyrektor otwarcie deklaruje brak jakiegokolwiek zainteresowania środowiskiem. Zgodnie z najbardziej prawdopodobnym wyjaśnieniem tego faktu wyrażenie „aprobuje szkodzenie” w języku polskim — odmiennie niż w języku angielskim — może być rozumiane jako „toleruje szkodzenie”⁴. Aprobata nie jest wówczas rozumiana jako postawa moralna w sensie bycia za czymś lub przeciw czemuś, lecz raczej jako przekonanie na temat czegoś. Tolerując szkodzenie, dyrektor nie musi przyjmować żadnej postawy (nie musi być ani zwolennikiem, ani przeciwnikiem szkodzenia). Wymaga się od niego jedynie zajęcia stanowiska, niekoniecznie związanego z jakimś stanem chcenia lub pragnienia. Badani, którzy wybierają tę odpowiedź, rozumieją prawdopodobnie aprobatę w sposób wychodzący poza opozycję aprobata—dezaprobata i zrównują przychylną postawę wobec czegoś z postawą tolerancyjną. Można więc przypuszczać, że wybierają tę odpowiedź również z intencją przypisania dyrektorowi postawy w pewien sposób indyferentnej. Zauważmy jednak, że nawet ignorując odpowiedzi tej części badanych, wciąż uzyskujemy jasne wyniki.

Druga hipoteza badawcza została natomiast sfalsyfikowana. Nie widać istotnej asymetrii między oceną postawy dyrektora w przypadku szkodzenia a jej oceną w przypadku pomagania. Z punktu widzenia hipotezy wyjściowej odpowiedzi na drugie pytanie są całkowicie zaskakujące. Okazuje się bowiem, że wartość moralna oczekiwanych skutków ubocznych działania dyrektora nie wpływa na ocenę jego postawy. Respondenci dezaprobuje samą postawę dyrektora niezależnie od tego, czy oczekiwane konsekwencje jego działania są pozytywne, czy negatywne. Zarówno w sytuacji, gdy działanie dyrektora miało pomóc środowisku, jak i w sytuacji, gdy miało mu zaszkodzić, ocena jego postawy była równie negatywna.

Brak asymetrii w ocenie postawy mógłby oznaczać, że ocena moralna postawy dyrektora następuje niejako przed oceną moralną oczekiwanych konsekwencji czynu i przed przypisaniem mu intencjonalności, które to przypisanie pozostaje pod wpływem rozmaitych rodzajów ocen moralnych. Gdyby bowiem wszystkie oceny odbywały się jednocześnie i były częścią jednej ogólnej oceny moralnej, to współzależności między nimi powinny wywołać asymetrię również w ocenie postawy. Wyniki badania wskazują, że tak nie jest, dlatego warto rozważyć hipotezę następstwa czasowego.

⁴ Wyjaśnienie to zostało zaproponowane przez prof. Katarzynę Paprzycką podczas pierwszej prezentacji wyników badań na seminarium „Studium teoretyczne i eksperymentalne efektu Knobe’a i problemu Butlera” w Instytucie Filozofii UW w roku akademickim 2014/2015.

Inną możliwością jest przemyślenie związków pojęciowych między intencjonalnością a moralnością. Skoro bowiem asymetria w przypisywaniu intencjonalności nie przekłada się na moralną ocenę postaw, to w takim razie intencjonalność nie może być traktowana jako pojęcie psychologii potocznej bezpośrednio związane z oceną moralną ujętą jako całościowy proces. Wyniki badania wskazują, że ocena intencjonalności nie wpływa na ocenę postaw, a więc niewykluczone, że brak tego wpływu ujawnia się również w innych obszarach oceny moralnej. Dlatego też intencjonalność jest być może związana nie tyle z ogólną oceną moralną danego działania, ile z niektórymi elementami tej oceny.

3. REWIZJA HIPOTEZY MORALNEJ

Pierwsza hipoteza badawcza została potwierdzona: niezależnie od scenariusza badani przypisują dyrektorowi taką samą postawę względem środowiska. Druga hipoteza została sfalsyfikowana: niezależnie od scenariusza badani jednakowo oceniają postawę dyrektora względem środowiska — piętnują ją mimo różnic między scenariuszami pod względem oczekiwanych konsekwencji działania. Oznacza to, że efekt Knobe'a nie wynika z przypisywania dyrektorowi odmiennych postaw względem środowiska ani też z tego, że w jednym scenariuszu jego postawa jest piętnowana, a w drugim nie. Jakże więc konsekwencje tych wyników dla całości rozważań nad efektem Knobe'a?

Chciałbym przedstawić trzy hipotezy które mogą stanowić punkt wyjścia dalszych badań. Po pierwsze, nasuwa się hipoteza, że przypisywanie intencjonalności służy wyrażeniu moralnej nagany dla samych oczekiwanych skutków działania. Skoro asymetria w przypisaniu intencjonalności nie ma odzwierciedlenia w ocenie postawy, to intencjonalność zawiera w sobie ocenę moralną innego elementu niż postawa działającego podmiotu. Najbardziej prawdopodobnym kandydatem wydają się tu oczekiwane konsekwencje podejmowanych przez podmiot działań.

Po drugie — ze względu na podobieństwo w sposobie oceniania postawy dyrektora w obydwu scenariuszach — można przypuszczać, że ocena postawy jest elementem oceny moralnej, który z łatwością daje się rozważać w oderwaniu od innych aspektów danej sytuacji. Respondenci nie mają problemu z wyizolowaniem samej postawy dyrektora i dokonaniem jej oceny, bez względu na inne czynniki, biorące udział w ocenie moralnej. Możliwość takiego wyizolowania pokazywałaby, że pojęcie postawy jest stosunkowo jasne i łatwo stosowalne przy dokonywaniu oceny moralnej.

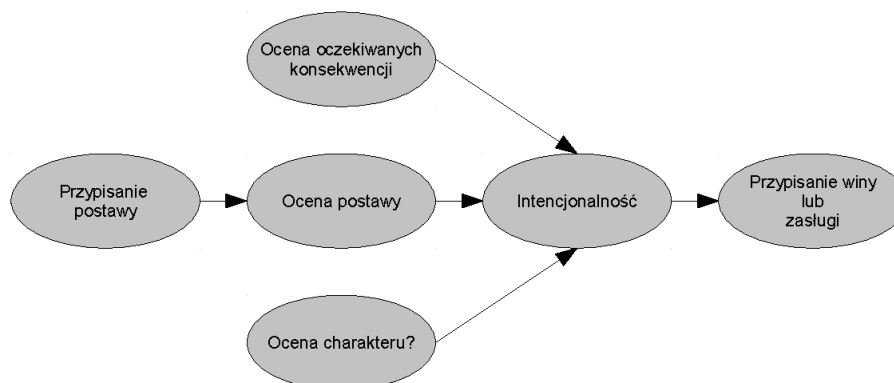
Po trzecie, wydaje się, że ludzie dokonują pewnej oceny moralnej, *zanim* dokonają przypisania intencjonalności. Przypisanie postawy moralnej nie wymaga znajomości oczekiwanych lub rzeczywistych skutków działania: wymaga jedynie oszacowania, czy podmiot działania jest za czymś, czy przeciw czemuś. Z kolei ocena tej postawy, jak wskazują badania, nie jest zależna od oczekiwanych skutków podjętego działania. Dlatego też intencjonalność można potraktować jako termin szerszy, obejmujący nie tylko działanie, lecz także jego konsekwencje. Przypisanie intencjonalności mogłoby zatem być rozumiane jako *rezultat* całej rozbudowanej oceny moralnej, w której poszczególne elementy różnią się co do swego wpływu na ocenę intencjonalności.

Porównywanie asymetrii efektu Knobe'a do asymetrii w ocenie poszczególnych elementów oceny moralnej mogłoby wykazać związki pojęciowe między intencjonalnością a terminami moralnymi. W dalszych badaniach należałoby więc sprawdzać, które elementy oceny moralnej mają największy wpływ na asymetryczne przypisanie intencjonalności. Być może, intencjonalność jest związana wyłącznie z oceną oczekiwanych konsekwencji albo wyłącznie z oceną charakteru działającego podmiotu, albo wyłącznie z oceną zgodności z etyką zawodową⁵. Niewykluczone też, że jest kombinacją kilku tych obszarów. Wymagałoby to wypracowania nowej definicji intencjonalności.

Możliwe jest także, że badania tego typu sfalsyfikują w ogóle hipotezę moralną. Jeżeli bowiem wyniki badań dotyczących wszystkich elementów oceny moralnej będą analogiczne, tj. asymetria nie będzie widoczna w stopniu zbliżonym do efektu Knobe'a, to należałoby wówczas przypuścić, że przypisywanie intencjonalności ma niewielki związek z oceną moralną.

Wyniki przedstawione w tym artykule pozwalają zatem nieco zmodyfikować schemat Knobe'a i Mendlowa, wyznaczając jednocześnie dalsze kierunki badań i nakreślając nowe możliwości interpretacyjne. Schemat ten wyglądałby następująco:

⁵ Jak zauważył recenzent, etyka zawodowa jest elementem z innego poziomu niż ocena konsekwencji lub ocena charakteru, dlatego też ich zestawienie może wydawać się nieuprawnione. Kierunek badań, który proponuję, ma jednak na celu jedynie zbadanie związku między intencjonalnością a poszczególnymi elementami, które podpadają pod szeroką kategorię oceny moralnej, niezależnie od tego, do którego poziomu lub podkategorii należą.



Rysunek 2. Rola intencjonalności w psychologii potocznej w świetle przedstawionych badań

Głównym celem dalszych badań powinno być drobiazgowo rozbięcie elementu „dobre czy złe” (por. Rysunek 1) na konkretniejsze elementy, możliwe do bezpośredniego, eksperymentalnego zbadania. W ten sposób ich wpływ na intencjonalność będzie mógł zostać dostrzeżony i oceniony pod względem istotności i intensywności.

BIBLIOGRAFIA

- Adams F., Steadman A. (2004), *Intentional Action in Ordinary Language. Core Concept or Pragmatic Understanding?*, „Analysis” 64(282), 173-181.
- Anscombe G. E. (1963), *Intention*, Oxford: Blackwell.
- Davidson D. (1963), *Actions, Reasons, and Causes*, „Journal of Philosophy” 60(23), 685-700.
- Dębska A. (2013), *Wnioskowanie na temat intencjonalności działania w ujęciu filozofii eksperymentalnej*, „Filozofia Nauki” 21(3) [83], 143-155.
- Harman G. (1976), *Practical Reasoning*, „The Review of Metaphysics” 29(3), 431-463.
- Hindriks F. (2008), *Intentional Action and the Praise-Blame Asymmetry*, „The Philosophical Quarterly” 58(233), 630-641.
- Knobe J. (2003a), *Intentional Action and Side Effects in Ordinary Language*, „Analysis” 63(3), 190-194.
- Knobe J. (2003b), *Intentional Action in Folk Psychology. An Experimental Investigation*, „Philosophical Psychology” 16(2), 309-324.
- 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. (2006), *The Concept of Intentional Action. A Case Study in the Uses of Folk Psychology*, „Philosophical Studies” 130(2), 203-231.

- Knobe J. (2010), *Person as Scientist, Person as Moralist*, „Behavioral and Brain Sciences”, 33(4), 315-329.
- 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.
- Nadelhoffer T. (2004), *On Praise, Side Effects, and Folk Ascription of Intentionality*, „Journal of Theoretical and Philosophical Psychology” 24(2), 196-213.
- Paprzycka K. (2014), *Rozwiązanie problemu Butlera i wyjaśnienie efektu Knobe’a*, „Filozofia Nauki” 22(2) [86], 73-96.
- Perry R. B. (1926), *General Theory of Value*, New York, NY: Longmans, Green.
- Satris S. (1987), *Ethical Emotivism*, Dordrecht: Martinus Nijhoff.
- Stevenson Ch. L. (1963), *Facts and Values*, New Heaven, CT: Yale University Press.
- Stevenson Ch. L. (1975), *Istota sporów etycznych [w:] Metaetyka*, I. Lazari-Pawłowska (red.), Warszawa: Państwowe Wydawnictwo Naukowe, 113-123.
- Wright J. C., Bengson J. (2009), *Asymmetries in Judgments of Responsibility and Intentional Action*, „Mind & Language”, 24(1), 24-50.