

JOHN MARTIN FISCHER

ZDOLNOŚĆ REAGOWANIA NA RACJE A ODPOWIEDZIALNOŚĆ MORALNA*

Odróżniamy istoty, którym można zasadnie przypisywać odpowiedzialność moralną za ich działania, od istot, którym takiej odpowiedzialności przypisywać nie można. Z kolei pośród działań moralnie odpowiedzialnego sprawcy rozróżniamy między działaniami, za które odpowiada on moralnie, a tymi, za które nie ponosi odpowiedzialności.

Sprawca jest moralnie odpowiedzialny za działanie w takim stopniu, w jakim inni, odpowiadając na to działanie, mogą racjonalnie zająć wobec niego określone postawy i podjąć określone czynności. Chodzi tu o takie nastawienia, jak uraza, oburzenie, szacunek i wdzięczność, oraz takie czynności, jak pochwała i nagana moralna, nagradzanie i karanie¹. Przy takim podejściu sprawcę można uznać za odpowiedni obiekt oceny moralnej — pochwały albo nagany — nawet wówczas, gdy w rzeczywistości nie zasługuje on na nie. Sprawca może na przykład odpowiadać moralnie za czyn, który jest moralnie „neutralny”. Teoria *odpowiedzialności moralnej* określa warunki, pod jakimi uznajemy, że ktoś jest *odpowiednim kandydatem* do po-

Prof. JOHN MARTIN FISCHER — profesor filozofii na University of California, Riverside; adres do korespondencji: Department of Philosophy, 900 University Ave., Riverside, CA 92521; e-mail: john.fischer@ucr.edu; ORCID: <https://orcid.org/0000-0002-1685-4743>.

* Przekład na podstawie: „Responsiveness and Moral Responsibility”, w: *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, red. Ferdinand Schoeman (Cambridge: Cambridge University Press, 1987), 81–106; przedruk w: John Martin FISCHER, *My Way: Essays on Moral Responsibility* (Oxford: Cambridge University Press, 2006), 63–83. Przekład za zgodą Autora.

Oryginał jest opatrzony następującym przypisem: „Wiele skorzystałem z uwag do wcześniejszej wersji tego tekstu poczynionych przez Sarahę Buss, Anthony’ego Bruecknera i Ferdinanda Schoemana. Skorzystałem również z możliwości odczytania wersji tego tekstu w Birbeck College, University of London”.

¹ Strawson nazywa takie postawy związane z odpowiedzialnością moralną „nastawieniami reaktywnymi” — STRAWSON 1962, [przekład polski: STRAWSON 2019].

chwały lub nagany z uwagi na swoje zachowanie. Taką teorię należy uzupełnić o teorię moralną określającą, których moralnie odpowiedzialnych sprawców (i w jakim wymiarze) *powinno się* chwalić lub ganić za ich działania. Obie teorie są oczywiście ważne, ale w tym eseju zajmę się tą pierwszą, czyli teorią, która wyjaśnia stosowność wspomnianych postaw i czynności.

Przedstawiam tu zaledwie szkic takiej teorii. Wymaga ona dokładniejszego opracowania i mocniejszej obrony. Ufam jednak, że taki szkic wystarczy, by pokazać, iż podejście to warto rozwinąć. Przedstawiona przeze mnie teoria nie jest z pewnością radykalnie nowa ani całkowicie różna od swoich poprzedniczek². Liczę jednak, że uda mi się doprecyzować ją w taki sposób, by uniknęła niektórych zarzutów zgłaszanych wobec podobnych stanowisk, i że zdołam wyprowadzić z niej kilka konsekwencji, które dotychczas nie zostały zauważone.

SZKIC TEORII ODPOWIEDZIALNOŚCI MORALNEJ

Teoria odpowiedzialności moralnej powinna uchwycić nasze intuicyjne sądy na temat niekontrowersyjnych przypadków. Zakładam, że w dużej mierze zgadzamy się co do pewnych przypadków, w których sprawcy można racjonalnie przypisać odpowiedzialność moralną za to, co robi, i co do pewnych przypadków, w których uczynić tego nie można. Przemysłane opinie na temat takich sytuacji stanowią ważne dane, które teoria odpowiedzialności moralnej ma za zadanie objaśnić. Aby wypracować zasadę, która mogłaby stanowić uzasadnienie naszych opinii, warto rozważyć najpierw te przypadki, w których skłonni jesteśmy sądzić, że sprawcy nie można zasadnie przypisać odpowiedzialności moralnej.

Wyobraźmy sobie człowieka poddanego hipnozie. Hipnotyzer wywołał w owym człowieku chęć, by na dźwięk telefonu uderzył osobę, która znajduje się w pobliżu. O ile człowiek ten nie wyraził zgody na takie hipnotyczne sugestie (podał się hipnozie na przykład po to, by rzucić palenie), przypisanie mu moralnej odpowiedzialności za to, że uderzył przyjaciela w nos, gdy zadzwonił telefon, wydaje się pozbawione sensu.

Przypuśćmy z kolei, że jakiś złoczyńca uzyskał dostęp do telewizora, który jest własnością Smitha, i tak go przeprogramował, iż jest w stanie poddać

² Niektóre współczesne wersje podobnych teorii można znaleźć w: MACINTYRE 1957; GLOVER 1970; FINGARETTE 1972; NEELY 1974; GERT i DUGGAN 1979; DAVIS 1979; LEVIN 1979; NOZICK 1981; DENNETT 1984. Znakomity przegląd wybranych aspektów tych stanowisk zob. SHATZ 1985.

Smitha wyszukanej formie oddziaływania podprogowego. Pod wpływem wielokrotnych oddziaływań tego rodzaju Smith morduje swojego sąsiada. Jest oczywiste, że Smithowi nie można przypisać odpowiedzialności moralnej za ów pożałowania godny czyn, z uwagi na przyczynową historię tego czynu.

Podobne odczucia budzą w nas również działania wywoływane na wiele innych sposobów. Sprawcom, których czyny są wynikiem skutecznych form prania mózgu czy indoktrynacji, wpływu silnych narkotyków albo manipulacji dokonanych bezpośrednio na mózgu, nie można sensownie przypisać odpowiedzialności moralnej za ich działania. Wyobraźmy sobie na przykład, że w przyszłości neurofizjologowie będą w stanie ustalić, które obszary mózgu należy poddać manipulacji, by wywołać określone decyzje czy działania. Jeśli pobudzą oni elektronicznie odpowiednie obszary mózgu Jonesa, sprawiając w ten sposób, że zapobiegnie on napadowi rabunkowemu, nie będzie można sensownie uznać, iż Jones odpowiada moralnie za swoje zachowanie. Nie będzie zasługą Jonesa, że udaremnił napad.

Podobnie, jeśli odkrywamy, że jakieś zachowanie jest wynikiem poważnego uszkodzenia mózgu albo neurologicznego zaburzenia, nie przypisujemy sprawcy moralnej odpowiedzialności za to zachowanie. Również pewne rodzaje zaburzeń psychicznych — takie jak skrajne odmiany fobii — mogą niekiedy wykluczać przypisanie sprawcy odpowiedzialności moralnej za jego zachowania.

Wielu ludzi sądzi, że mogą istnieć autentycznie „nieodparte” bodźce psychiczne. Gdyby tak było w istocie, także one mogłyby powodować zachowanie, za które sprawcy nie można przypisać odpowiedzialności. Narkomani (w pewnych warunkach) mogą działać pod wpływem dosłownie nieodpartych bodźców, zatem nie moglibyśmy obciążać ich odpowiedzialnością moralną za działanie pod wpływem tych pragnień (zwłaszcza gdy uważamy, że nie odpowiadają moralnie za swoje uzależnienie).

Także pewne rodzaje gróźb (i być może propozycji) z użyciem przymusu wykluczają moralną odpowiedzialność. Kasjer w banku, który słyszy, że zostanie zastrzelony, jeśli nie odda pieniędzy, może odczuwać przemożne i nieodparte pragnienie, by spełnić to żądanie. O ile działa pod wpływem takiego impulsu, można sensownie domniemywać, że nie odpowiada za swoje czyny³.

Nie ulega wątpliwości, że przypisując odpowiedzialność moralną, bierzemy pod uwagę przyczynową historię danego działania. Osoby poddane

³ Można przeciwstawić tego kasjera innemu, który w dokładanie takich samych okolicznościach nie ma nieodpartego impulsu, by postąpić zgodnie z żądaniem. Takiemu kasjerowi można przypisać moralną odpowiedzialność za przekazanie pieniędzy (co nie oznacza, że jest on *godny potępienia*).

manipulacji są jak marionetki i trudno przypisać im zasługę lub winę. Rozumiemy intuicyjnie, że niektóre czynniki prowadzące do zachowania podważają odpowiedzialność.

Przypadki, w których działa czynnik podważający odpowiedzialność, można zestawić z tymi, w których mechanizm namysłu działa „normalnie”, bez zakłóceń. Kiedy zastanawiasz się, czy przeznaczyć pięć procent twojej pensji na organizację charytatywną United Way, ważąc racje za i przeciw, a twojej decyzji o przekazaniu pieniędzy nie wywołała hipnoza, pranie mózgu, bezpośrednia manipulacja, bodźce psychotyczne itp., to uważamy, że można cię zasadnie pochwalić za ten uczynek. O ile nie dostrzegamy żadnego czynnika podważającego odpowiedzialność, który wpływałby na twoją decyzję czy działanie, jesteśmy skłonni uważać cię za osobę moralnie odpowiedzialną.

Można by sądzić, że istnieje stosunkowo oczywisty sposób odróżnienia jasnych przypadków odpowiedzialności moralnej od jasnych przypadków jej braku. Wydaje się, że w przypadkach, w których sprawca jest moralnie odpowiedzialny za działanie, ma wolność postąpienia inaczej, natomiast w przypadkach braku odpowiedzialności jest pozbawiony takiej wolności. Inaczej mówiąc, faktyczne działanie czynnika, który intuicyjnie podważa odpowiedzialność moralną, podważa ją dlatego, że wyklucza wolność postąpienia inaczej.

Można to ująć następująco. Gdy sprawca jest, powiedzmy, zahipnotyzowany, wówczas nie reaguje on w odpowiedni sposób na racje. Ze względu na stan hipnozy zachowałby się tak samo niezależnie od tego, jakie racje wchodziłyby w grę. Przypuśćmy tak jak wcześniej, że pod hipnozą otrzymał instrukcję, by na dźwięk telefonu uderzyć znajdującą się w pobliżu osobę. Z tego względu po usłyszeniu dzwonka telefonu uderzyłby ją, nawet gdyby miał niezwykle silne racje, by tego nie robić. Sprawca nie reaguje na racje — jego zachowanie pozostałoby takie samo, bez względu na to, jakie racje wchodziłyby w grę.

Kiedy z kolei ludzki mechanizm namysłu działa normalnie, bez zakłóceń, przyjmujemy, że sprawca *reaguje* na racje. Gdy postanowiasz przekazać pieniądze na United Way of America, uważamy, że nie zrobiłbyś tego, gdybyś odkrył, że organizacja ta dopuściła się grubego nadużycia finansowego. Naturalnie i rozsądnie jest więc przyjąć, że różnica między moralnie odpowiedzialnymi sprawcami, a tymi, którzy nie są odpowiedzialni, polega na „zdolności reagowania na racje”.

Uważam jednak, że istnieją przypadki, w których sprawcy można przypisać odpowiedzialność moralną za określone działanie, nawet gdy nie mógł

postąpić inaczej (i nie jest „zdolny reagować na racje”)⁴. Zilustrujmy to przykładem. Wyobraźmy sobie, że jakiś złoczyńca zainstalował w mózgu Browna urządzenie, które pozwala mu śledzić aktywność psychiczną Browna i ingerować w nią bez żadnych ograniczeń. Ów złoczyńca może elektronicznie sterować mózgiem Browna za pomocą „pilota”, wywołując w ten sposób określone decyzje. Wyobraźmy sobie, że może on również sprawić, iż Brown będzie działał zgodnie z tak wywołanymi decyzjami. Przypuśćmy teraz, że Brown jest bliski zamordowania swego sąsiada, czego właśnie pragnie złoczyńca. Inaczej mówiąc, wyobraźmy sobie, że wspomniane urządzenie jedynie monitoruje aktywność mózgu Browna, ale nie przyczynia się w żaden sposób do podjęcia przez niego faktycznej decyzji i działania. Brown dokonuje namysłu i działa dokładnie tak, jak gdyby w jego mózgu nie umieszczono żadnego urządzenia. Wyobraźmy sobie również, że gdyby Brown zaczął podejmować decyzję o niezamordowaniu sąsiada, urządzenie uruchomiłoby się i za jego sprawą Brown postanowiłby popełnić morderstwo (i popełniłby je). Mamy tu sytuację, w której możemy przypisać sprawcy odpowiedzialność moralną za wykonanie działania, mimo że nie mógł on postąpić inaczej⁵. Nazwijmy ten przykład przykładem frankfurtofskim.

W przykładzie frankfurtofskim faktyczne zdarzenia przebiegają tak, że zasadne jest przypisanie odpowiedzialności moralnej, chociaż alternatywny scenariusz (a być może cały wachlarz takich scenariuszy) przebiega w sposób wykluczający odpowiedzialność. W faktycznej sekwencji zdarzeń nie pojawia się żaden czynnik podważający odpowiedzialność, choć pojawia się on w scenariuszu alternatywnym. Takie przykłady przemawiają na rzecz bardziej wyrafinowanej teorii odpowiedzialności moralnej, odwołującej się do faktycznej sekwencji zdarzeń. W ramach takiej teorii rozróżniamy między mechanizmami działającymi w faktycznej sekwencji zdarzeń a tymi, które działają w alternatywnej sekwencji (czy alternatywnych sekwencjach).

⁴ John Locke w *Essay Concerning Human Understanding*, ks. II, rozdz. 12 par. 8–11 [LOCKE 1690/1894; polski przekład: LOCKE 1955] przedstawił interesujący przykład człowieka dobrowolnie pozostającego w pokoju, który — o czym ów człowiek nie wie — jest zamknięty. Wiele przykładów sprawców, którzy są moralnie odpowiedzialni za swoje działania, choć nie mogli postąpić inaczej, można znaleźć w FRANKFURT 1969, 829–39 [przekład polski: FRANKFURT 1997b]. Zob. również FISCHER 1982.

⁵ Napisaną z werwą i wielce interesującą krytykę takiego opisu tego przypadku przedstawia Peter van Inwagen (INVAGEN 1978; przedruk w: INWAGEN 1983). Choć tu nie miejsce, by wchodzić w szczegóły tej debaty, bronię w niej tezy, że są przypadki, gdy sprawca jest moralnie odpowiedzialny za wykonanie działania, chociaż nie mógł postąpić inaczej. Piszę o tym w FISCHER 1982.

W przykładzie frankfurtowskim faktycznie działający rodzaj mechanizmu reaguje na racje, w przeciwieństwie do mechanizmu, który zadziałałby w alternatywnym scenariuszu⁶. W przypadku omówionym wyżej działanie Browna jest następstwem normalnej władzy praktycznego rozumowania, o której możemy sensownie przyjąć, że reaguje na racje. Za to w alternatywnym scenariuszu aktywny byłby inny rodzaj mechanizmu, który bezpośrednio stymulowałby mózg Browna. Ów mechanizm nie reagowałby na racje. Zatem mechanizm działający w faktycznej sekwencji zdarzeń może reagować na racje, chociaż sam *sprawca* nie reaguje na nie. (*Brown* nie mógł postąpić inaczej, niż postąpił).

Proponuję więc, by wprowadzić następujący, bardziej precyzyjny sposób odróżniania stosunkowo jasnych przypadków moralnej odpowiedzialności od tych, w których tej odpowiedzialności brak. Sprawca odpowiada moralnie za działanie, o ile mechanizm, który faktycznie do niego prowadzi, reaguje na racje. W przypadku mechanizmu, który nie reaguje na racje, sprawca nie może postąpić inaczej, niż postępuje. Jednakże sprawca, który nie może postąpić inaczej, niż postępuje, może działać w oparciu o mechanizm reagujący na racje, a w związku z tym można przypisać mu odpowiedzialność za to, co robi.

Podałem pewne przykłady, w których jest intuicyjnie jasne, że sprawcy nie można przypisać moralnej odpowiedzialności za to, co uczynił, i inne przykłady, w których jest intuicyjnie jasne, że można mu taką odpowiedzialność przypisać. Zaproponowałem też zasadę pozwalającą odróżnić te dwa typy przypadków. Zasada ta odwołuje się do dwóch względów: zdolności reagowania na racje oraz rozróżnienia między mechanizmami działającymi w faktycznej sekwencji zdarzeń i mechanizmami działającymi w alternatywnej sekwencji. Sformułowanie tej zasady jest jednak nieprecyzyjne i skrótowe. Wyjaśnię ją teraz dokładniej, zaczynając od pojęcia zdolności reagowania na racje.

ZDOLNOŚĆ REAGOWANIA NA RACJE

Chciałbym omówić dwa typy zdolności reagowania na racje: mocny i słaby. Zacznijmy od mocnego. Mamy z nim do czynienia wtedy, gdy pewien mechanizm rodzaju *R* faktycznie prowadzi do działania, i gdyby istniała wystarczająca racja, by postąpić inaczej, a *R* byłby aktywny, sprawca rozpoznałby tę rację, podjął decyzję, że postąpi inaczej, i tak by postąpił. W celu ustalenia, czy dany mechanizm ma zdolność mocnego reagowania

⁶ Ten sposób opisanie przykładów frankfurtowskich zawdzięczam Sydneyowi Shoemakerowi.

na racje, należy spytać, co by się stało, gdyby istniała wystarczająca racja, by sprawca postąpił inaczej, i aktywny byłby mechanizm z faktycznej sekwencji zdarzeń. W takiej sytuacji sprawca musi spełnić trzy warunki: uznać, że racje są wystarczające, podjąć decyzję zgodną z tymi racjami i postąpić zgodnie z podjętą decyzją. Odpowiednio mogą wystąpić co najmniej trzy rodzaje usterek w „alternatywnej sekwencji” zdarzeń: usterka w powiązaniu między racjami, jakie faktycznie istnieją, a racjami, jakie rozpoznaje sprawca, usterka w powiązaniu między racjami a decyzją sprawcy, oraz usterka w powiązaniu między decyzją a działaniem.

Pierwszy rodzaj usterki to *niewrażliwość* na racje. Występuje ona u niektórych psychopatów z urojeniami⁷. Drugi rodzaj to *brak reakcji* — niereagowanie w odpowiedni sposób na [własne] przekonania. Brak takiej reakcji charakteryzuje niektórych neurotyków z natręctwami lub lękami⁸. Istnieje wreszcie usterka w przełożeniu decyzji na działanie; ta usterka to rodzaj bezsilności. Gdyby w alternatywnej sekwencji nie wystąpiła żadna z tych usterek (i działałby faktyczny rodzaj mechanizmu), to faktycznie działający mechanizm cechowałaby zdolność mocnego reagowania na racje. Istniejące racje ściśle dokładnie odpowiadałyby racjom uznanym przez sprawcę, racje sprawcy dokładnie odpowiadałyby jego wyborowi, a wybór dokładnie odpowiadałby działaniu. Działanie sprawcy *dokładnie* odwzorowywałoby racje⁹.

Uważam, że gdy źródłem działania jest mechanizm zdolny mocno reagować na racje, spełniony jest wystarczający warunek odpowiedzialności moralnej; nie sądzę jednak, by zdolność mocnego reagowanie na racje była jej warunkiem koniecznym. Aby się o tym przekonać, wyobraźmy sobie, że wskutek poprawnego funkcjonowania normalnej ludzkiej władzy rozumowania praktycznego, postanawiam dziś wieczór pójść na mecz koszykówki (i idę na ten mecz), oraz że mam po temu wystarczającą rację. Przypuśćmy jednak, że wykazałbym „słabą wolę”, gdyby istniała wystarczająca racja, by *nie* iść na mecz. Załóżmy, że taką wystarczającą racją byłby nieprzekraczalny termin oddania ważnego maszynopisu (termin, którego nie mógłbym dotrzymać, gdybym udał się na mecz). Mimo to wybrałbym pójście na mecz, w pełni zdając sobie sprawę, że mam wystarczającą rację, by zostać w domu

⁷ Piszą o tym GERT i DUGGAN 1979.

⁸ Ibid.

⁹ Robert Nozick wymaga takiego typu dokładnego odwzorowania działania na wartości, gdy mówi o „śledzenia wartości”. Zob. NOZICK 1981. Tak więc pod tym względem wprowadzone przez Nozicka pojęcie śledzenia wartości odpowiada mocnej zdolności reagowania na racje. Nozick twierdzi, że sprawca, który śledzi wartości, wykazuje się pewną cnotą moralną, ale nie twierdzi, że śledzenie wartości jest koniecznym warunkiem odpowiedzialności moralnej.

i zająć się pracą. Wydaje mi się, że moje faktyczne pójście na mecz koszykówki jest działaniem wolnym i można rozsądnie przypisać mi za nie odpowiedzialność moralną, mimo że jego źródłem w faktycznej sekwencji jest mechanizm, który — z powodu mojej słabej woli — nie ma mocnej zdolności reagowania na racje.

Pójście na mecz koszykówki można wiarygodnie potraktować jako działanie moralnie neutralne; zgodnie z przyjętym tu rozumieniem odpowiedzialności można moralnie odpowiadać za działanie, nawet gdy nie zasługuje ono ani na pochwałę, ani na naganę. Zjawisko słabej woli stanowi również problem dla intuicyjnie jasnych przypadków odpowiedzialności moralnej za działania *godne pochwały*. Załóżmy, że spędzam popołudnie pracując na rzecz organizacji charytatywnej United Way (a moja decyzja i działanie są wynikiem mechanizmu, który, intuicyjne, stanowi podstawę odpowiedzialności). Wyobraźmy sobie, że gdybym miał wystarczającą rację, aby powstrzymać się od tego działania, byłby nią (raz jeszcze) nieprzekraczalny termin złożenia tekstu do publikacji. Powiedzmy, że poświęciłbym mój czas działalności charytatywnej, nawet gdybym miał wystarczającą rację, by tego nie robić. W tym wypadku, jak się zdaje, jestem moralnie odpowiedzialny i zasługuję na pochwałę za to, co robię, mimo że faktyczny mechanizm nie jest zdolny do mocnego reagowania na racje.

Nie ulega ponadto wątpliwości, że mocna zdolność reagowania na racje nie może być koniecznym warunkiem odpowiedzialności moralnej za moralnie naganne i/lub nieroztropne działania. Przypuśćmy, że kradnę ze sklepu książkę, wiedząc doskonale, iż jest to moralnie niesłuszne, oraz że zostanę przyłapany, a zatem moje działanie jest też nieroztropne. Faktyczna sekwencja może mimo to, w intuicyjnej ocenie, stanowić podstawę odpowiedzialności; mogą nie działać w niej żadne czynniki, które intuicyjnie podważają odpowiedzialność moralną. (Oczywiście zakładam, że mogą istnieć autentyczne przypadki działań będących wyrazem słabej woli, a zarazem wolnych i pozwalających na przypisanie odpowiedzialności sprawcy). Oto więc przypadek, w którym jestem moralnie odpowiedzialny za kradzież książki, mimo że mechanizm w faktycznej sekwencji nie ma zdolności mocnego reagowania na racje. Istnieje faktycznie wystarczająca racja (zarówno moralna, jak roztropnościowa), by postąpić inaczej, a mimo to dopuszczam się kradzieży książki.

Wszystkie trzy podane wyżej przypadki są kłopotliwe dla twierdzenia, że odpowiedzialność moralna wymaga mocnej zdolności reagowania na racje. Takie reagowanie może być wystarczające i konieczne dla pewnego rodzaju

pochwały — powiązanie działań z wzorami wartości w taki sposób, że człowiek mocno reaguje na racje, to dowód wielkiej cnoty. Oczywiście nie każdy moralnie odpowiedzialny sprawca jest moralnie godny pochwały (czy choćby maksymalnie roztropny). Uważam, że do odpowiedzialności moralnej wystarczy mniej dokładna odpowiedniość między racjami i działaniem, mianowicie „słaba zdolność reagowania na racje”.

W wypadku mocnej zdolności reagowania na racje pytamy, co by się stało, gdyby istniała wystarczająca racja, by postąpić inaczej (przy zachowaniu faktycznego typu mechanizmu). Taka zdolność kieruje nas do alternatywnego scenariusza, w którym istnieje wystarczająca racja, by sprawca postąpił inaczej (i działa faktyczny mechanizm), który *najbardziej przypomina* faktyczną sytuację. Mówiąc w języku światów możliwych, w wypadku mocnej zdolności reagowania na racje w grę wchodzi te spośród możliwych światów, w których sprawca ma dostateczną rację, by postąpić inaczej (i w których działa faktyczny typ mechanizmu), które są *najbardziej zbliżone* do aktualnego świata. (Być może istnieje tylko jeden taki świat, być może jest ich więcej. W wypadku słabej zdolności reagowania na racje, przeciwnie, musi istnieć *jakiś* możliwy świat, w którym jest dostateczna racja, by postąpić inaczej, aktywny jest faktyczny mechanizm sprawcy i sprawca postępuje inaczej. Ten możliwy świat (a może ich być więcej niż jeden) nie musi być tym światem, w którym sprawca ma wystarczającą rację, by postąpić inaczej (i działa faktyczny mechanizm), który jest *najbardziej zbliżony* do aktualnego świata¹⁰.

Raz jeszcze weźmy pod uwagę moją decyzję, by pójść na mecz koszykówki. W tej sytuacji, jeśli miałbym wystarczającą rację, by postąpić inaczej, byłby nią ostateczny termin złożenia tekstu do publikacji; wykazałbym się w tej sytuacji słabą wolą i mimo wszystko poszedł na mecz. Z pewnością jednak istnieje *jakiś* scenariusz, w którym działa faktyczny mechanizm, mam dostateczną rację, by nie pójść na mecz, i nie idę na mecz. Przypuśćmy na przykład, że powiedziano mi, iż za bilet wstępu na mecz będę musiał zapłacić tysiąc dolarów. Chociaż mam skłonność, by w pewnych okolicznościach okazywać słabą wolę, to istnieją warunki, w których zareagowałbym odpowiednio na wystarczające racje. Chodzi mianowicie o warunki, w których istniejące racje są zdecydowanie *mocniejsze* niż te, które istniałyby, gdybym miał wystarczającą rację, by postąpić inaczej.

¹⁰ Przyjmuję tu założenie, że w możliwych światach istotnych dla słabej zdolności reagowania na racje przez mechanizm z faktycznej sekwencji obowiązują te same *prawa przyrody*, co w aktualnym świecie.

Weźmy z kolei pod uwagę godną pochwały pracę, jaką po południu wykonałem na rzecz organizacji charytatywnej United Way. Choć zrobiłbym to i tak, nawet gdybym miał nieprzekraczalny termin oddania tekstu do druku, to z pewnością *nie* pracowałbym dla United Way, gdyby miało to doprowadzić do utraty przeze mnie stałego zatrudnienia. A zatem faktyczny mechanizm, który prowadzi do mojego działania, słabo reaguje na racje. Podobnie, kiedy sprawca w sposób niesłuszny (i nieroztropny) kradnie książkę (tj. istnieje wystarczająca racja, by tego nie czynić), faktyczny mechanizm, który stoi za tym działaniem, mógłby reagować na przynajmniej pewną logicznie możliwą zachętę, by powstrzymać się od kradzieży. O ile ów mechanizm tak reaguje na racje, o tyle można sprawcy w sposób właściwy przypisać moralną odpowiedzialność za kradzież książki. Nawet sprawca, który działa wbrew dobrym racjom, może reagować na *pewne* racje.

Uważam, że w faktycznej sekwencji mechanizm *musi* [przynajmniej] słabo reagować na racje, jeśli mamy przypisać sprawcy odpowiedzialność moralną. Gdyby (zakładając działanie faktycznego rodzaju mechanizmu) sprawca ukradł książkę, nawet wiedząc, że w ten sposób doprowadzi do śmierci własnej i swojej rodziny, to wydaje się, że faktyczny mechanizm wykluczałby przypisanie mu odpowiedzialności moralnej za to działanie.

Sprawca, którego działanie jest wytworem mechanizmu mocno reagującego na racje, zasługuje na pochwałę; jego zachowanie dokładnie odwzorowuje wartości. Odpowiedzialność moralna jednak wymaga tylko mechanizmu słabo reagującego na racje. W ramach proponowanego przeze mnie podejścia faktyczna irracjonalność da się pogodzić z odpowiedzialnością moralną (tak jak należy). Opisany przez Dostojewskiego człowiek z podziemia jest być może przykładem osoby faktycznie irracjonalnej, a mimo to moralnie odpowiedzialnej. Podobnie pewne rodzaje hipotetycznej irracjonalności są zgodne z odpowiedzialnością moralną; skłonność do słabości woli nie musi być dowodem żadnej usterki w faktycznym mechanizmie prowadzącym do działania. Odpowiedzialność moralna wymaga *pewnego* związku między racją a działaniem, ale ów związek może być dość luźny¹¹.

¹¹ Ferdinand Schoeman zwrócił moją uwagę na pewien rodzaj przykładu, który zagraża memu twierdzeniu, że słaba zdolność reagowania na racje wystarcza do moralnej odpowiedzialności. Wyobraźmy sobie szaleńca. Osoba ta popełnia barbarzyński akt, na przykład szablą zabija wiele osób na promie Staten Island. Założmy też, że osoba ta uśmierciłaby tych ludzi w każdych możliwych warunkach z jednym wyjątkiem: powstrzymałaby się, gdyby wiedziała, że jest piątek, czyli święto. Intuicyjnie rzecz biorąc, osoba ta jest wysoce irracjonalna i nie powinno się uważać jej za odpowiedzialną moralnie, a jednak wydaje się ona spełniać warunek działania w oparciu o mechanizm mający słabą zdolność reagowania na racje. Mamy tu do czynienia ze słabą zdol-

Rozróżniłem w tej części dwa rodzaje zdolności reagowania na racje. Dowodziłem, że sprawca jest moralnie odpowiedzialny za działanie, o ile jego źródłem jest mechanizm, który [przynajmniej] słabo reaguje na racje. W następnej części omówię analogię między tą teorią odpowiedzialności moralnej a stanowiącą jej odpowiednik teorią wiedzy. Owa analogia pomoże nam lepiej zrozumieć to, że podstawą odpowiedzialności moralnej jest faktyczna sekwencja. W kolejnej części doprecyzuję teorię odpowiedzialności moralnej, nadając bardziej określony kształt newralgicznej idei „rodzaju mechanizmu prowadzącego do działania”.

WIEDZA A ODPOWIEDZIALNOŚĆ

Zarysowałem model odpowiedzialności moralnej oparty na faktycznej sekwencji. Wedle niego sprawca może odpowiadać moralnie za wykonanie działania, mimo że nie jest w stanie postąpić inaczej. Wystarczy, że w faktycznej sekwencji mechanizm odpowiednio reaguje na racje. Zachodzi analogia między tego rodzaju teorią odpowiedzialności moralnej a modelem wiedzy opartym na faktycznej sekwencji. Zgodnie z tym modelem podmiot może mieć wiedzę o określonym sądzie logicznym, nawet jeśli nie ma odpowiedniej zdolności rozróżniania; wystarczy, że mechanizm w faktycznej sekwencji jest odpowiednio wrażliwy na prawdziwość.

Nie ulega wątpliwości, iż osoba wie, że p , tylko wtedy, gdy jest przekonana, że p , oraz p jest prawdziwy; spełnienie jednak tych dwóch warunków z pewnością nie wystarcza. Istnieją rozmaite strategie nałożenia [na wiedzę] dodatkowych warunków¹². Jedno z „eksternalistycznych” podejść głosi, iż przekonanie, że p , musi być „rzetelnym wskaźnikiem” prawdziwości p lub że musi ono „śledzić” prawdziwość p . Z grubsza można

nością reagowania na racje z uwagi na to, że choć sprawca nie reaguje na szereg istotnych racji, reaguje on na pewną dziwną rację.

Zdaję sobie sprawę z tego, że przykład ten stanowi problem dla teorii odpowiedzialności, jaką tu przedstawiam. Widzę aktualnie dwie możliwe odpowiedzi. Po pierwsze, można byliby twierdzić, że w tego typu wypadkach w alternatywnej sekwencji (w której sprawca reaguje na racje) działałby inny mechanizm niż w faktycznej sekwencji. Alternatywnie, można byliby jakoś ograniczyć racje istotne dla słabej zdolności reagowania na racje. Mam nadzieję omówić takie przykłady i wypracować adekwatne stanowisko w tej sprawie w przyszłości.

¹² Z grubsza można wyróżnić „internalistyczne” i „eksternalistyczne” koncepcje wiedzy. Internalista wymaga, by sprawca miał pewien rodzaj *uzasadnienia* dla swojego przekonania. Eksternalista rezygnuje z doprecyzowywania rodzajów uzasadnień, wymaga za to pewnych rodzajów związku przyczynowego między znanym faktem a przekonaniem sprawcy.

by powiedzieć, iż osoba ma wiedzę, że p , tylko wtedy, gdy spełnia zarazem następujące warunki: (1) nie miałaby przekonania, iż p , gdyby p nie był prawdziwy, i (2) w rozmaitych okolicznościach, w których p byłby prawdziwy, posiadałaby przekonanie, że p . Mówimy tu o przekonaniach danej osoby w obrębie światów, które są stosunkowo podobne do rzeczywistego — zarówno tych, w których p jest prawdziwy, jak i tych, w których p jest fałszywy¹³.

Przypuśćmy, że jadąc samochodem widzisz coś, co przypomina stojącą na polu stodołę, dochodzisz do wniosku, że istotnie jest to stodoła, a twój wniosek jest prawdziwy. Choć nic ci o tym nie wiadomo, gdyby nie było tam zwykłej stodoły, demoniczny farmer postawiłby w jej miejsce replikę wykonaną z papier *mâché*. W tej sytuacji masz prawdziwe przekonanie, że widzisz zwykłą stodołę, ale twoje przekonanie nie „śledzi” prawdy: nawet gdyby na polu nie było stodoły, pozostałabyś przekonany, że stodoła tam jest. Nie masz więc zdolności rozróżniania, która, jak się zdaje, stanowi warunek konieczny wiedzy.

Porównajmy ten przypadek z innym, w którym widzisz w supermarkecie banan i dochodzisz do wniosku, że na półce istotnie leży banan. Zakładamy, że nie istnieje demoniczny kierownik supermarketu gotowy wyprowadzić cię w pole i że gdyby na półce nie było banana, nie miałabyś przekonania, o którym mowa. W tym wypadku twoje przekonanie zapewne śledzi prawdę i można powiedzieć, że wiesz, iż na półce leży banan. Co więcej, posiadasz taką wiedzę, mimo że logicznie możliwy *jest* scenariusz, w którym demoniczny kierownik sklepu kładzie na półce plastikową imitację banana, a ty, jak poprzednio, dochodzisz do wniosku, że to banan. Dla wiedzy istotne są te scenariusze obejmujące prawdziwość p , które *najbardziej przypominają* świat aktualny; to, że istnieją odleglejsze możliwości, w których sąd p jest fałszywy, nie ma znaczenia¹⁴.

Opisane przypadki mogłyby sugerować, iż podmiot ma wiedzę, że p , tylko wtedy, gdy potrafi odróżnić warunki, jakie zachodziłyby, gdyby p był prawdziwy, od tych, które zachodziłyby, gdyby p był fałszywy. Rozważmy jednak następujące przykłady (autorstwa Nozicka):

¹³ Przedstawiam tu oczywiście zaledwie szkic teorii wiedzy. Co więcej, nie zakładam, że jest to bezsprzecznie *prawdziwa* koncepcja. Wskazuję jedynie na analogię między moim podejściem do odpowiedzialności moralnej a eksteralistyczną koncepcją wiedzy. Zaprezentowane tu podejście do wiedzy opiera się między innymi na stanowisku Dretskego i Nozicka. Zob. DRETSKE 1971 oraz NOZICK 1981, 167–198. Nozick także omawia analogię między odpowiedzialnością moralną a wiedzą.

¹⁴ Nozick twierdzi, że fakt ten pozwala obalić pewien rodzaj epistemologicznego sceptycyzmu. Zob. NOZICK 1981, 197–247.

Babcia widzi, że jej wnuk ma się dobrze, gdy ten ją odwiedza; gdyby jednak był chory czy martwy, ludzie, chcąc oszczędzić jej przykrości, powiedzieliby jej, że wnuk ma się dobrze. Nie oznacza to wszak, że babcia, widząc swego wnuka, nie wie, iż jej wnuk jest zdrow (albo w każdym razie jakoś się trzyma). (Nozick 1981, 179)

S ma przekonanie, że w pewnym budynku mieści się teatr i filharmonia. Bywała w nim na sztukach i koncertach. [...] Gdyby jednak ów budynek nie był teatrem, mieściłby się w nim reaktor jądrowy, który (założmy) tak zmieniłby okoliczne powietrze, że każdy, kto znalazłby się poblizu, popadałaby w letarg, dostawał mdłości i rezygnował z próby kupienia biletu. Sfabrykowana przez rząd historyjka mówiłaby, że budynek jest teatrem; historyjki tej rząd mógłby być pewien, bo bez zabezpieczeń medycznych żadna osoba nie mogłaby przejść przez obszar [wywołujący] mdłości, by odkryć prawdę. Założmy, że wszyscy uwierzyliby w rządową historyjkę i uznali, że budynek, który widzą (ale tylko z pewnej odległości), jest teatrem. (Ibid., 180–181)¹⁵

Przykłady te są epistemologicznymi odpowiednikami przykładów w stylu frankfurtowskim, w których sprawca jest moralnie odpowiedzialny za wykonanie działania, chociaż nie mógł postąpić inaczej. Podmiot wie, że p , chociaż brak mu odpowiedniej zdolności rozróżniania. Tak jak w wypadku odpowiedzialności moralnej zrezygnowaliśmy z wymogu, by sprawca był zdolny reagować na racje, na rzecz wymogu, by mechanizm miał taką zdolność, tak i w wypadku prawdy poprzestajemy na wymogu, by mechanizm był wrażliwy na prawdę.

Jak zauważa Nozick, można mieć przekonanie, że p , uzyskane za pośrednictwem mechanizmu wrażliwego na prawdę, i w ten sposób wiedzieć, że p , chociaż w alternatywnym scenariuszu (czy scenariuszach) działałby niewrażliwy mechanizm. Poszukujemy więc teorii wiedzy opartej na faktycznej sekwencji, tak jak poszukujemy teorii odpowiedzialności opartej na faktycznej sekwencji. Powinniśmy odróżnić mechanizm z faktycznej sekwencji od mechanizmu z sekwencji alternatywnej i skupić się na własnościach tego pierwszego. Choć jednak zachodzi mocna analogia między przedstawionymi teoriami odpowiedzialności i wiedzy, między odpowiedzialnością a wiedzą zachodzą dwie istotne różnice, które teraz wskażę.

Po pierwsze, w myśl podanej wyżej teorii odpowiedzialności, jeśli sprawca działa w oparciu o mechanizm typu M , to warunkiem jego odpowiedzialności moralnej jest istnienie *jakiegoś* możliwego scenariusza, w którym M działa, sprawca ma wystarczającą rację, by postąpić inaczej, i postępuje inaczej. Ów możliwy scenariusz nie musi być tym, który zostałby zrealizowany, gdyby M działał i sprawca miał wystarczającą rację, by postąpić inaczej. Innymi słowy,

¹⁵ Nozick przypisuje ten przykład Avishai Margalitowi.

scenariusz (albo zestaw scenariuszy), w którym działa mechanizm typu M , a sprawca ma dostateczną rację, by postąpić inaczej, istotny dla przypisania odpowiedzialności, nie musi być scenariuszem *najbardziej podobnym* do faktycznego. W podanej wyżej teorii wiedzy jest wręcz przeciwnie: jeśli sprawca ma przekonanie, że p , uzyskane za pośrednictwem mechanizmu typu M , to jest konieczne, iż gdyby aktywny był mechanizm typu M i p był fałszywy, to warunkiem posiadania przez podmiot wiedzy, że p , byłoby posiadanie przezeń przekonania, iż p jest fałszywy.

Z grubsza biorąc, logiczne możliwości istotne dla przypisania odpowiedzialności moralnej mogą być odleglejsze niż te, które bierzemy pod uwagę przypisując wiedzę. Dlatego uważam, że związek między racjami a działaniem, którego wymaga odpowiedzialność moralna, jest „luźniejszy” niż konieczny do wiedzy związek między prawdą a przekonaniem. Oczywiście nie kłóci się to z twierdzeniem, że zarówno wiedza, jak odpowiedzialność moralna są pojęciami opartymi na faktycznej sekwencji; chodzi tylko o to, że wrażliwość na prawdę w faktycznej sekwencji jest zdefiniowana „ściślej” (tj. w kategoriach „bliższych” możliwości) niż zdolność reagowania na racje w faktycznej sekwencji.

Między odpowiedzialnością moralną a wiedzą zachodzi jeszcze, jak sądzę, druga różnica. Twierdziłem, że tak jak odpowiedzialność moralna nie wymaga wolności, by postąpić inaczej, tak wiedza nie wymaga zdolności różnicowania; w wypadku odpowiedzialności wystarczy zdolność reagowania na racje, a w wypadku wiedzy — wrażliwość na prawdę. Tak więc oba pojęcia dotyczą faktycznej sekwencji. Chcę jednak wskazać na mocniejszy sens, w którym odpowiedzialność moralna, w odróżnieniu od wiedzy, zależy tylko od faktycznej sekwencji.

Twierdzą mianowicie, że odpowiedzialność moralna sprawcy za działanie superwenuje na faktycznych, fizycznych relacjach przyczynowych, które prowadzą do działania, natomiast wiedza podmiotu, że p , *nie* superwenuje na faktycznych, fizycznych relacjach przyczynowych, które prowadzą do przekonania, że p . Wyjaśnię najpierw twierdzenie o superwencji w odniesieniu do odpowiedzialności moralnej. Wydaje mi się niemożliwe, by istniały przypadki, w których mamy dwóch sprawców, którzy wykonują działanie tego samego typu w następstwie dokładnie tego samego rodzaju faktycznej sekwencji przyczynowej, a zarazem tylko jeden z nich jest moralnie odpowiedzialny za to działanie. Różnice w przypisaniu odpowiedzialności muszą być pochodną różnic w faktycznych czynnikach fizycznych prowadzących do działania; same różnice w alternatywnych scenariuszach nie przekładają

się na różnice w przypisaniu odpowiedzialności. Inaczej mówiąc, różnice w przypisaniu odpowiedzialności muszą wynikać z różnic w faktycznych, a nie tylko „możliwych”, historiach działań.

Przypuśćmy, że ty i ja heroicznie skaczemy do jeziora, by ocalić tonącego pływaka, a faktyczny przebieg zdarzeń jest pod istotnymi względami podobny — z tym tylko wyjątkiem, że ty mógłbyś postąpić inaczej, a ja nie. (Nie mógłbym postąpić inaczej ze względu na istnienie w moim mózgu mechanizmu, który wywołałby w nim decyzję ocalenia pływaka, w razie gdybym skłaniał się ku innej decyzji). W takiej mierze, w jakiej faktyczne sekwencje fizyczne prowadzące do działania są identyczne, nasza odpowiedzialność moralna jest taka sama.

Oto jednak epistemologiczny przykład Nozicka:

Rozważmy przypadek, studenta, który, gdy jego zajęcia z filozofii zostają odwołane, zwykle wraca do swojego pokoju i zażywa halucynogeny. Jednym ze złudzeń, jakim czasem ulega pod ich wpływem, jest poczucie bycie na zajęciach z filozofii. Czy kiedy ów student faktycznie jest na zajęciach z filozofii, wie, że na nich jest? Myślę, że nie wie, ponieważ gdyby nie był na zajęciach, nadal mógłby mieć przekonanie, że na nich jest. [...] Dwóch studentów na tych zajęciach mogłoby być w tej samej faktycznej sytuacji, mając (z grubsza) te same pobudzenia wzrokowe i słuchowe, a jednak pierwszy wiedziałby, że przebywa na zajęciach, a drugi nie, ponieważ są inaczej usytuowani „modalnie” — są o nich prawdziwe inne twierdzenia w trybie warunkowym. (NOZICK 1981, 191)

Dokładnie te same faktyczne czynniki fizyczne wywołują u tych dwóch studentów przekonanie, że przebywają na zajęciach. Mimo to jeden z nich nie wie, że jest na zajęciach; gdyby nie przebywał na zajęciach (i posłużył się metodą introspekcji, którą faktycznie się posłużył), to (za sprawą halucynogenów) nadal byłby (lub mógłby być) przekonany, że jest na zajęciach. Drugi student — który nie ma zwyczaju brać halucynogenów — wie, że przebywa na zajęciach. Wiedza zatem, w odróżnieniu od odpowiedzialności moralnej, nie superwenuje na faktycznych fizycznych faktach.

Twierdziłem wyżej, że zachodzi podobieństwo między odpowiedzialnością moralną a wiedzą: jak zdolność reagowania na racje faktycznego mechanizmu prowadzącego do działania wystarczy do odpowiedzialności, tak wrażliwość na prawdę faktycznego mechanizmu prowadzącego do przekonania wystarczy do wiedzy. W jaki sposób uzgodnić to twierdzenie o podobieństwie z twierdzeniem, że przypisanie odpowiedzialności moralnej, w przeciwieństwie do przypisania wiedzy, superwenuje na faktycznych fizycznych czynnikach przyczynowych? Myślę, że odpowiedź kryje się w naszym intuicyjnym sposobie

jednostkowienia „mechanizmów”. W teorii działania jesteśmy skłonni rozróżniać je bardziej szczegółowo niż w epistemologii.

W wypadku pierwszego studenta uznajemy, że właściwym mechanizmem prowadzącym do przekonania jest „introspekcja”. Oczywiście ten sam rodzaj mechanizmu działałby, gdyby student zażył halucynogen. Przy takim „szerokim” sposobie jednostkowienia, okazuje się, że mechanizm prowadzący do przekonania jednego studenta *nie* jest wrażliwy na prawdę, a drugiego *jest*.

Z kolei w wypadku, w którym („samodzielnie”) ratuję tonące dziecko, nasuwa się przypuszczenie, że gdybym został pobudzony [do tego działania] przez naukowców, mielibyśmy do czynienia z *innym* mechanizmem niż ten, który działa faktycznie. Podobnie, gdyby wstrzyknięto mi środek wywołujący nieodparte pragnienie ocalenia tonącego pływaka, ten rodzaj mechanizmu *różniłby się od* faktycznego. Przy tym „wąskim” sposobie jednostkowienia okazuje się, że mechanizm, który prowadzi do ocalenia przeze mnie dziecka, *jest* mechanizmem reagującym na racje (tak samo jak twój).

Asymetryczność superwencji można uzgodnić z symetrycznością opartej na faktycznej sekwencji wiedzy i odpowiedzialności moralnej. Asymetryczność superwencji jest następstwem intuicyjnie naturalnej skłonności do jednostkowienia mechanizmów tworzenia przekonań mniej dokładnie niż mechanizmów prowadzących do działania¹⁶.

MECHANIZMY

Wyżej postawiłem tezę, że osoba jest moralnie odpowiedzialna za wykonanie działania, jeśli mechanizm, który faktycznie do niego prowadzi, reaguje na racje. Teza ta wymaga doprecyzowania, ponieważ w danej sytuacji mogą operować różne mechanizmy. Który z nich jest istotny dla przypisania odpowiedzialności?

Przypuśćmy, że (w normalny sposób) namyślam się, czy przeznaczyć pięć procent pensji na United Way, postanawiam to zrobić i działam zgodnie z podjętą decyzją. Moglibyśmy tak uszczegółowić ten przykład, by stał się intuicyjnie wzorcowym przykładem sytuacji, w której odpowiadam moralnie

¹⁶ Pozostawiłem tu kluczowe pojęcie „tych samych mechanizmów” w postaci niezmiernie ogólnej. Nie ulega wątpliwości, że pojęcie to stanowi źródło niepokojących problemów w epistemologii. Omówienie niektórych z nich można znaleźć w SHOPE 1984, 29–48. Z pewnością podobne problemy mogą się pojawić w teorii działania. Na potrzeby przypisywania odpowiedzialności moralnej polegam na intuicyjnym sposobie jednostkowienia rodzajów mechanizmów prowadzących do działania. Dalsza obrona teorii, którą tu w zarysie przedstawiam, wymagałaby dokładniejszego omówienia jednostkowienia mechanizmów.

za swoje działanie. Przyjrzyjmy się jednak mechanizmowi, który tu faktycznie działa, czyli „namysłowi poprzedzającemu przeznaczenie pięciu procent pensji na United Way”. Gdyby działał *ten* rodzaj mechanizmu, to oddałbym pięć procent mojej wypłaty na United Way w dowolnym logicznie możliwym scenariuszu. Ten zatem rodzaj faktycznie działającego mechanizmu *nie* reaguje na racje.

Taki jednak mechanizm jak „namysł poprzedzający przeznaczenie pięciu procent pensji na United Way” nie ma znaczenia dla przypisania odpowiedzialności moralnej. Nie jest to bowiem mechanizm „czasowo wewnętrzny”. Działanie mechanizmu czasowo zewnętrznego czy „relacyjnego” zawiera już w sobie działanie, którego niby jest przyczyną.

Zwróćmy uwagę, że z działania mechanizmu w rodzaju „namysłu poprzedzającego przeznaczenie pięciu procent pensji na United Way” *wynika*, iż sprawca przeznaczy tę kwotę na United Way. W tym sensie wspomniany mechanizm z góry obejmuje dane działanie: jego funkcjonowanie implikuje, że dane działanie zostanie wykonane. Mechanizm zatem (w myśl proponowanej tu teorii) istotny dla przypisania odpowiedzialności moralnej musi być „czasowo wewnętrzny” czy „nierelacyjny” w następującym sensie: jeśli mechanizm *M* prowadzi do działania *X*, *M* jest istotny dla odpowiedzialności moralnej sprawcy za wykonanie *X* tylko wtedy, gdy z funkcjonowania *M* nie wynika, że *X* zajdzie. Uważam ten wymóg za intuicyjnie naturalny i bezsporny. Oczywiście jak na razie mamy tylko konieczny warunek bycia istotnym mechanizmem; różne mechanizmy prowadzące do działania mogą być czasowo wewnętrzne. Który z nich jest „tym właściwym”?

Nie dysponuję teorią pozwalającą odpowiedzieć na to pytanie w sposób ogólny. Przedstawiona wyżej teoria zakłada po prostu, że na potrzeby oceny odpowiedzialności moralnej dla każdego działania zostaje odpowiednio wybrany pewien intuicyjnie naturalny mechanizm jako „ten właściwy”.

To założenie nie wydaje mi się problematyczne. W razie gdyby budziło ono zastrzeżenia, warto odnotować, że podstawowa teoria może się bez niego obejść. W podanej wyżej wersji teoria ta głosi, że sprawca odpowiada moralnie za wykonanie działania, o ile (istotny, czasowo wewnętrzny) mechanizm prowadzący do działania reaguje na racje. Można by zamiast tego powiedzieć, że sprawca jest moralnie odpowiedzialny za działanie, o ile nie istnieje faktycznie działający czasowo wewnętrzny mechanizm prowadzący do działania, który nie reaguje na racje. To alternatywne sformułowanie pozwala obejść wymóg wskazania jednego mechanizmu jako „tego właściwego”. W dalszym ciągu pozostaję przy pierwszym sformułowaniu, ale główne tezy powinny zachować ważność również przy sformułowaniu alternatywnym.

Zastosujmy teraz tę teorię do kilku przypadków. Intuicyjnie sądzimy, że nieodparta chęć może stanowić rodzaj przymusu psychicznego i wykluczać odpowiedzialność moralną. Wyobraźmy sobie, że Jim ma dosłownie nieodpartą chęć, by zażyć pewien narkotyk, i tak czyni. Jaki dokładnie jest istotny mechanizm, który prowadzi do tego działania? Zauważmy, że nie może to być mechanizm „namysłu obejmującego nieodpartą chęć zażycia narkotyku”, ponieważ nie jest on czasowo wewnętrzny. Jego działanie zawiera już to, że Jim zażyje narkotyku. Rozważmy zatem mechanizm „namysłu obejmującego nieodpartą pragnienie”. Mechanizm ten *jest* czasowo wewnętrzny i reaguje na racje. Istnieje możliwy scenariusz, w którym Jim działa w oparciu o tego typu mechanizm i powstrzymuje się od wzięcia narkotyku. W tym scenariuszu Jim ma nieodpartą chęć *powstrzymania się* od zażycia narkotyku. Te rozważania pokazują, że ani „namysł obejmujący nieodpartą pragnienie zażycia narkotyku”, ani „namysł obejmujący nieodpartą pragnienie” nie są istotnymi mechanizmami (o ile teoria odpowiedzialności ma współgrać z naszymi intuicyjnymi sądami).

Kiedy Jim działa zgodnie z nieodpartą chęcią zażycia narkotyku, w jego centralnym układzie nerwowym zachodzi pewien fizyczny proces rodzaju *F*. Gdy osoba podlega tego rodzaju procesowi, mówimy, że chęć jest dosłownie nieodparta. Powodem, dla którego intuicyjnie nie przypisujemy Jimowi odpowiedzialności moralnej za zażycie narkotyku, jest — jak sądzę — to, że istotny rodzaj mechanizmu prowadzący do zażycia przez niego narkotyku należy do fizycznego rodzaju *F*, a mechanizm rodzaju *F* nie reaguje na racje. Gdy sprawca działa zgodnie z dosłownie nieodpartą chęcią, podlega pewnemu rodzajowi faktycznego procesu fizycznego, który nie ma zdolności reagowania na racje, a to wyklucza odpowiedzialność moralną¹⁷.

Wróćmy do mojej tezy, że określone rodzaje „bezpośredniego manipulowania mózgiem” wykluczają odpowiedzialność moralną. Nie dotyczy to oczywiście wszystkich takich manipulacji. Przypuśćmy na przykład, że jakiś naukowiec manipuluje jedną komórką na obrzeżach mojego mózgu. Nie musi to wykluczać mojej odpowiedzialności, o ile tego rodzaju proces fizyczny może reagować na racje. Dopiero gdy naukowcy ingerują w mózg i manipulują nim tak, że nie reaguje on na racje, podważają moralną odpowiedzialność sprawcy za działanie¹⁸.

¹⁷ To twierdzenie, w postaci, w jakiej zostało tu sformułowane, opiera się na intuicji, że fizyczny proces *F* jest istotnym mechanizmem. Alternatywnie, można byłoby po prostu wskazać, że w wypadku Jima faktycznie *działa* mechanizm (typu *F*), który jest czasowo wewnętrzny i nie reaguje na racje.

¹⁸ Daniel Dennett pisze tak: „Możliwość zaburzenia pracy systemu intencjonalnego czy innej ingerencji w ten system prowadzi do wielu kłopotliwych pytań dotyczących stopnia odpowie-

Podobnie nie wszystkie formy podprogowych oddziaływań, hipnozy, prania mózgu itd. kłócą się z odpowiedzialnością moralną. Do wykluczenia odpowiedzialności dochodzi jedynie wówczas, gdy wytworzone w wyniku wspomnianych ingerencji mechanizmy fizyczne nie mają zdolności reagowania na racje. Teoria, która łączy odpowiedzialność moralną z reagowaniem na racje w faktycznej sekwencji, może więc ułatwić wyjaśnienie naszych intuicyjnych rozróżnień między oddziaływaniami przyczynowymi, które są spójne z odpowiedzialnością moralną, a tymi, które ją wykluczają.

Weźmy pod uwagę klasę obron sądowych, które można by nazwać obronami „typu Twinkie”. Taka obrona głosi, że sprawcy nie należy karać, ponieważ spożył zbyt wiele śmieciowego jedzenia (co uszkodziło jego zdolności poznawcze itd.). Zgodnie z przyjętym tu podejściem pytanie, czy sprawca powinien być ukarany, rozpada się na dwa: (1) Czy sprawca jest moralnie odpowiedzialny (tj. racjonalnie może podlegać karze)? oraz (2) jeśli tak, to jaki wymiar kary jest odpowiedni? Przedstawiona przeze mnie teoria odpowiedzialności moralnej pozwala odpowiedzieć twierdząco na pierwsze pytanie w typowej sprawie „typu Twinkie”.

Nawet jeśli ktoś jadł wyłącznie potrawy śmieciowe, trudno wyobrazić sobie, że w efekcie doszło do procesu biologicznego, który nawet słabo nie reaguje na racje. Co najwyżej taki proces może nie mieć mocnej zdolności reagowania na racje, ale taka zdolność *nie* jest koniecznym warunkiem odpowiedzialności moralnej. Oburzenie, jakie wywołuje w nas sugestia, że ktoś, kto żywi się śmieciowym jedzeniem, nie podlega odpowiedzialności moralnej, może mieć dwa źródła. Może być reakcją bądź na filozoficznie błędny wymóg posiadania mocnej, a nie słabej, zdolności reagowania na racje, bądź na trudny do zaakceptowania pomysł, że spożywanie śmieciowego jedzenia daje w efekcie mechanizm, który nawet słabo nie reaguje na racje.

Moja teoria odpowiedzialności uzasadnia więc intuicję, że pozwani typu Twinkie odpowiadają moralnie za swoje czyny. Osobne pytanie dotyczy odpowiedniego wymiaru kary; warto jednak zauważyć, że z teorii odpowiedzialności *nie* wynika, by sprawca działający w oparciu o mechanizm, który słabo (a nie mocno) reaguje na racje, zasługiwał na *łagodniejszy* wymiar kary niż sprawca, który działa w oparciu o mechanizm mocno reagujący na racje. Zasada ta jest opcjonalnym składnikiem pełnej teorii kary.

działalności w sytuacji, gdy miała miejsce manipulacja. Ogólnie, jesteśmy zwolnieni z odpowiedzialności, jeśli padliśmy ofiarą manipulacji; nie ma jednak jednej zasady niewinności z racji manipulacji” (DENNETT 1978, zwłaszcza 248). Moja propozycja pozwala odróżnić manipulację podważającą odpowiedzialność od manipulacji, która jest z nią spójna.

UWAGI DOTYCZĄCE WYMIARU CZASOWEGO

Przejdę teraz do omówienia pewnej trudności, przed jaką staje opracowywana przeze mnie teoria odpowiedzialności. Wymusi to doprecyzowanie tej teorii. Przypuśćmy, że Maks (który lubi wypić, ale nie jest uzależniony) upija się na przyjęciu prawie do nieprzytomności. W tym stanie wsiada do samochodu i próbuje wrócić do domu. Niestety na przejściu dla pieszych potrąca dziecko. Mimo że w faktycznej sekwencji mechanizm prowadzący do potrącenia dziecka zapewne nie ma zdolności reagowania na racje, możemy mieć poczucie, że Maks odpowiada moralnie za to, co zrobił.

Przykład ten należy do [ogólniejszej] klasy przypadków. W tych przypadkach sprawca działa w chwili t_1 w oparciu o reagujący na racje mechanizm, który sprawia, że w t_2 działa on według mechanizmu *niereagującego* na racje. I tak Maks powinien był wiedzieć, że upicie się na przyjęciu doprowadzi do sytuacji, w której pokieruje samochodem w stanie niepozwalającym mu reagować na racje. Można więc pociągnąć go do odpowiedzialności za jego działanie w chwili t_2 ze względu na istnienie odpowiedniego rodzaju mechanizmu reagującego na racje we wcześniejszej chwili t_1 . Gdy sprawca działa w chwili t_1 w oparciu o mechanizm reagujący na racje i powinien wiedzieć, że to działanie w pewnej późniejszej chwili t_2 doprowadzi go do działania w oparciu o mechanizm niereagujący na racje, można przypisać mu odpowiedzialność moralną za działanie w t_2 . Zgodnie z omawianą tu teorią, odpowiedzialność moralna za działanie w t wymaga zatem faktycznego operowania reagującego na racje mechanizmu w t lub w odpowiednio wcześniejszym czasie. (Dla uproszczenia pomijam dalej indeksy czasowe).

Ktoś mógłby wykształcić w sobie gotowość do cnotliwego działania w określonych sytuacjach. Niewykluczone nawet, że działając cnotliwie miałby tak silną motywację, iż mechanizm nie reagowałby na racje. O ile jednak do pielęgnowania cnoty doprowadziły go mechanizmy reagujące na racje, można przypisać mu odpowiedzialność za działanie. Jedynie wtedy, gdy w żadnym momencie procesu wiodącego do działania nie zadziałał mechanizm reagujący na racje, sprawcy nie można zasadnie pociągnąć do odpowiedzialności moralnej.

SEMIKOMPATYBILIZM

Przedstawiłem tu w dużym skrócie teorię odpowiedzialności moralnej. Podstawową ideę należałoby rozwinąć i wyjaśnić znacznie staranniej, to jednak, co powiedziałem, wystarczy, by wyprowadzić z tej teorii pewne konsekwencje. Twierdzę, że zarysowana teoria prowadzi do kompatybilizmu w kwestii relacji między odpowiedzialnością moralną a takimi doktrynami, jak przewidziana Boga i determinizm przyczynowy.

Zacznijmy od relacji między determinizmem przyczynowym a odpowiedzialnością moralną. Przedstawiona tu teoria pomaga nam uzgodnić determinizm przyczynowy z odpowiedzialnością moralną, nawet jeśli taki determinizm jest niezgodny z wolnością postąpienia inaczej. Argument za wzajemną niezgodnością determinizmu i wolności jest różny (i mocniejszy) od argumentu za niezgodnością determinizmu przyczynowego i odpowiedzialności moralnej.

Determinizm przyczynowy można zdefiniować następująco:

Determinizm przyczynowy to teza, że w dowolnej chwili pełny wykaz faktów o świecie w owej chwili, łącznie z pełnym wykazem praw przyrody, pociąga logicznie każdą prawdę dotyczącą tego, co wydarzy się później.

Możemy teraz podać „podstawowy argument” za niezgodnością determinizmu przyczynowego i wolności postąpienia inaczej. Jeśli obowiązuje taki determinizm, to (mówiąc w uproszczeniu) z połączenia przeszłości i praw przyrody wynika, że działam teraz tak, jak działałem. Jeśli więc jestem wolny postąpić inaczej, to mam albo władzę nad przeszłością, albo władzę nad prawami przyrody. Skoro jednak przeszłość i prawa przyrody są „ustalone” — na przykład nie mogę teraz postąpić tak, żeby przeszłość była inna, niż faktycznie była, to nie mam teraz wolności postąpić inaczej¹⁹.

Jest to oczywiście skrótowa prezentacja „podstawowego argumentu”; jego staranniejsze i szczegółowsze omówienie wykracza poza ramy tej prezentacji²⁰. Powinno być wszakże jasne, że kompatybilista w kwestii determinizmu przyczynowego i wolności postąpienia inaczej musi odrzucić albo ustaloną naturę przeszłości, albo ustaloną naturę praw przyrody. Inaczej mówiąc, musi powiedzieć, że w jakiejś chwili sprawca ma moc postąpić w taki sposób, żeby przeszłość była inna niż faktycznie była, lub że sprawca może

¹⁹ Niektóre współczesne rozwinięcia „podstawowego argumentu” na rzecz inkompatabilizmu można znaleźć w: GINET 1966; WIGGINS 1973; LAMB 1977; VAN INWAGEN 1975 oraz 1983, zwłaszcza 55–105.

²⁰ Omawiam ten argument w FISCHER 1983, 127–37; FISCHER 1986, 252–260; FISCHER 1988, 235–52. Jego klasyczne omówienie można znaleźć w LEWIS 1981, 113–121.

postąpić w taki sposób, by jakieś prawo przyrody, które faktycznie obowiązuje, nie obowiązywało²¹. Nawet jeśli te twierdzenia nie są jawnie fałszywe, to z pewnością trudno je zaakceptować.

Opracowane tu podejście do odpowiedzialności moralnej pozwala odzielić kompatybilizm w kwestii determinizmu przyczynowego i odpowiedzialności moralnej od kompatybilizmu w kwestii takiego determinizmu i wolności postąpienia inaczej. Podejście to głosi, że sprawcę można pociągnąć do odpowiedzialności moralnej za wykonanie działania, o ile mechanizm faktycznie prowadzący do tego działania ma zdolność reagowania na racje; sprawca nie musi być wolny postąpić inaczej. Jak wyjaśniam dalej, reagowanie na racje przez mechanizm w faktycznej sekwencji jest zgodne z determinizmem przyczynowym. Dlatego kompatybilista w kwestii determinizmu i odpowiedzialności moralnej może *przyjąć* ustaloną naturę przeszłości i praw przyrody. Nie musi akceptować nieatrakcyjnych twierdzeń, na które jest skazany kompatybilista w kwestii determinizmu przyczynowego i wolności postąpienia inaczej. Przyjęcie pod wpływem „podstawowego argumentu” inkompatabilizmu w kwestii determinizmu przyczynowego i wolności postąpienia inaczej, nie wymusza automatycznie przyjęcia inkompatabilizmu w kwestii takiego determinizmu i odpowiedzialności moralnej.

Teoria odpowiedzialności wymaga mechanizmów reagujących na racje. Aby mechanizm reagował na racje, możliwy musi być taki scenariusz, w którym działa ten sam rodzaj mechanizmu, a sprawca postępuje inaczej; identyczność rodzaju mechanizmu nie wymaga jednak identyczności w każdym, choćby najmniejszym, szczególe. Nasza intuicyjna koncepcja rodzaju mechanizmu prowadzącego do działania ani nasze sądy o jasnych przypadkach odpowiedzialności moralnej nie wskazują bynajmniej, że identyczność rodzaju mechanizmu implikuje identyczność na poziomie mikroszczegółów. Dlatego scenariusze istotne dla reagowania na racje przez mechanizm w faktycznej sekwencji mogą różnić się zarówno pod względem bodźców skłaniających sprawcę, by postąpić inaczej, jak i konkretnych szczegółów mechanizmu prowadzącego do działania. (Zauważmy, że jeśli obowiązuje determinizm przyczynowy i czynię *X*, jednym z faktycznie działających rodzajów mechanizmu jest mechanizm typu „przyczynowo zdeterminowany, by uczynić *X*”. Co jednak oczywiste, ten rodzaj mechanizmu nie ma znaczenia dla przypisywania odpowiedzialności moralnej, o ile nie jest czasowo wewnętrzny. A jeśli rodzaj „przyczynowo zdeterminowany” jest

²¹ Inne interesujące wyzwanie wobec pewnych sformułowań „podstawowego argumentu” można znaleźć w SLOTE 1982, 5–24.

czasowo wewnętrzny i jako taki może mieć znaczenie, to reaguje on na racje).

Jeśli istotnie obowiązuje determinizm przyczynowy, to każdy możliwy scenariusz (zakładający faktyczne prawa przyrody), w którym sprawca postępuje inaczej w chwili t , musi w czasie wcześniejszym niż t różnić się pod *pewnym* względem od faktycznego scenariusza. Teoria odpowiedzialności moralnej wymaga jedynie istnienia takich możliwych scenariuszy. Nie jest konieczne, by sprawca mógł urzeczywistnić taki scenariusz (tj. by w mocy sprawcy w t było postąpienie tak, że to, co przeszłe względem t , byłoby inne niż faktycznie było). Co więcej, istnienie wymaganego rodzaju scenariuszy jest zgodne z determinizmem przyczynowym.

Teoria odpowiedzialności moralnej oparta na zdolności reagowania na racje w aktualnej sekwencji daje nam w rezultacie „semikompatybilizm”: odpowiedzialność moralna jest zgodna z determinizmem przyczynowym, nawet jeśli taki determinizm wyklucza wolność postąpienia inaczej. Kompatybilizm w kwestii determinizmu i odpowiedzialności jest zgodny *zarówno* z kompatybilizmem, jak też z inkompatybilizmem (a także z agnostycyzmem) w kwestii determinizmu i wolności postąpienia inaczej²².

Inkompatybiliści często przywołują omówiony wyżej przykład demonicznych naukowców, którzy przeprowadzają bezpośrednie manipulacje na ludzkich mózgach, po czym pytają kompatybilistę: w jaki sposób tego rodzaju przypadek *różni się* od sytuacji, w której obowiązuje determinizm przyczynowy. Z pewnością zachodzi tu następujące podobieństwo: zarówno w przypadku manipulacji, jak i podlegania determinizmowi, by spowodować działanie, wystarczą warunki całkowicie „zewnętrzne” wobec sprawcy. Być może więc ani w jednym ani w drugim przypadku sprawca nie ma wolności postąpić inaczej. Jak jednak argumentowałem wyżej, wydaje się, że zachodzi kluczowa różnica między przypadkiem bezpośredniej manipulacji a „li tylko” determinacją przyczynową. W wypadku bezpośredniego manipulowania mózgiem proces prowadzący do działania prawdopodobnie nie reaguje na racje, podczas gdy fakt, że proces jest przyczynowo deterministyczny nie przesądza jeszcze, czy reaguje on na racje. Siłą wyzwania inkompatybilistycznego

²² Sądzę, że Frankfurt jest kompatybilistycznym semikompatybilistą. Ja natomiast jestem agnostycznym semikompatybilistą, a być może utajonym inkompatybilistycznym semikompatybilistą. W FISCHER 1982 wskazałem, że przykłady frankfurtofskie same w sobie nie dowodzą zgodności przyczynowego determinizmu i odpowiedzialności moralnej. Pozostawiają one zatem otwartą kwestię stanowiska „ultra-inkompatybilizmu”: przyczynowy determinizm jest niezgodny z odpowiedzialnością moralną, nawet jeśli odpowiedzialność moralna nie wymaga wolności postąpienia inaczej. Tu wolałem agnostyczny (albo być może inkompatybilistyczny) semikompatybilizm od agnostycznego (albo inkompatybilistycznego) ultrainkompatybilizmu.

jest, jak można przyjąć, wiarygodna idea, że w żadnym z obu przypadków sprawca nie ma wolności postąpienia inaczej, ale można na to wyzwanie odpowiedzieć, wskazując na różnicę w mechanizmach działających w faktycznej sekwencji.

Z tych samych względów odpowiedzialność moralna jest zgodna z boską przedwiedzą, nawet jeśli taka przedwiedza jest niezgodna z wolnością postąpienia inaczej. Przyjmijmy, że Bóg istnieje i dokładnie wiedział w przeszłości, jak zachowam się dziś. Jeśli mam wolność postąpić inaczej, to muszę być wolny tak postąpić, że przeszłość byłaby inna niż faktycznie była (tj. tak postąpić, żeby Bóg miał wtedy inne przekonanie o moim zachowaniu od tego, jakie faktycznie miał). Przeszłość jest już jednak ustalona, można więc wiarygodnie sądzić, że nie jestem wolny postąpić inaczej, jeśli Bóg istnieje.

Istnienie Boga da się jednak bez wątpienia pogodzić z działaniem mechanizmu, który reaguje na racje. Przekonanie Boga nie jest częścią mechanizmu prowadzącego do mojego działania (przy standardowym poglądzie na naturę Boga). To nie boskie przekonanie powoduje moje działanie; odwrotnie, moje działanie wyjaśnia boskie przekonanie. Istnieją zatem możliwe scenariusze, w których faktyczny mechanizm działa, wskutek czego postępuję inaczej. (W tych scenariuszach Bóg ma poprawne przekonanie, że postąpię inaczej, niż postępuję w aktualnym świecie). Ponownie argumenty na rzecz obu rodzajów inkompetybilizmu — w kwestii boskiej przedwiedzy i odpowiedzialności oraz w kwestii boskiej przedwiedzy i wolności postąpienia inaczej — są *różne*, teoria zaś oparta na zdolności reagowania na racje w faktycznej sekwencji daje nam semikompetybilizm²³.

²³ Próbowałem zarysować tu podejście, które próbuje uzgodnić odpowiedzialność moralną za działanie z przyczynowym determinizmem i przedwiedzą Boga. Moje podejście opiera się na twierdzeniu, że odpowiedzialność moralna za działanie nie wymaga wolności postąpienia inaczej. Gdzie indziej argumentowałam, że podczas gdy sprawca może być moralnie odpowiedzialny za wykonanie działania, chociaż nie mógł postąpić inaczej, to nie można przypisać mu odpowiedzialności za niewykonanie działania, którego nie mógł wykonać: FISCHER 1985–1986, 251–270. Jeśli ta „teza o asymetrii” jest prawdziwa, to nie uzgodniłem jeszcze odpowiedzialności moralnej za *zaniechania* (albo być może za „nie-zrobienia”) z przyczynowym determinizmem (i boską przedwiedzą).

Nie mam tu miejsca, by w pełni rozwinąć moją teorię odpowiedzialności za zaniechanie. Mogę jednak powiedzieć, że nawet jeśli sprawca nie jest odpowiedzialny za niezrobienie czegoś, czego nie mógł zrobić, to można mu jednak przypisać odpowiedzialność moralną za *coś* (na przykład za jakieś „pozytywne” działanie). Będzie on zatem zasługiwał na pochwałę lub naganę. Sądzę, że można rozwinąć taką teorię odpowiedzialności moralnej tak, by pogodzić przyczynowy determinizm (i boską przedwiedzę) z postawami moralnymi, o których sądzimy, że intuicyjnie są poprawne.

STRUKTURA A HISTORIA

W tej części chcę porównać moje podejście do odpowiedzialności moralnej z innymi teoriami odpowiedzialności, które można by nazwać teoriami opartymi na „dopasowaniu”. Moje podejście uwzględnia historię.

Rozważmy najpierw „hierarchiczny” model odpowiedzialności moralnej. Zgodnie z nim osoba odpowiada za działanie, o ile zachodzi dopasowanie między preferencją wyższego rzędu a preferencją pierwszego rzędu, która faktycznie skłania osobę do działania. W myśl jednej wersji tej teorii, zasugerowanej przez pewne uwagi Harry’ego Frankfurta, sprawca odpowiada za działanie, jeśli istnieje zgodność między jego „chcieniem drugiego rzędu” a „wołą” (pragnieniem pierwszego rzędu, które prowadzi go do działania)²⁴.

W innej wersji tej teorii, odpowiedzialność moralna za działanie wymaga dopasowania między „utożsamieniem się” a wołą²⁵. Zgodnie z sugestią Frankfurta, jednym ze sposobów utożsamienia się z pragnieniem pierwszego rzędu byłoby dojście do nienapotykanego oporu chcenia drugiego rzędu, by działać zgodnie z tym pragnieniem, w połączeniu z sądem, że dalszy namysł nie doprowadziłby do zmiany decyzji.

Problem z takimi hierarchicznymi teoriami „dopasowania” polega na tym, że niezależnie od tego, jak się je uściśla, dopasowanie można wytworzyć za pomocą mechanizmów podważających odpowiedzialność. Demoniczny neurofizjolog może wszak wytworzyć zgodność między różnymi psychicznymi elementami za pomocą jakiegoś rodzaju bezpośredniego elektro-nicznego oddziaływania, które nie reaguje na racje. Na tym właśnie, w moim przekonaniu, polega problem z teoriami hierarchicznego dopasowania: są one czysto strukturalne i ahistoryczne. Tymczasem liczy się również to, jaki rodzaj procesu prowadzi do działania. W szczególności, mechanizm prowadzący do działania musi reagować na racje.

Czysto strukturalny charakter mają również teorie dopasowania uwzględniające „wiele źródeł”. Zamiast hierarchii preferencji postulują one różne źródła preferencji. Zgodnie z jedną taką teorią, autorstwa Gary’ego Watsona, istnieją „preferencje wartościujące” (pochodzące od rozumu) oraz preferencje motywacyjne (WATSON 1975, 205–220). Posiłkując się teorią Watsona, można by powiedzieć, że sprawca jest moralnie odpowiedzialny za działanie, o ile

²⁴ FRANKFURT 1971, 5–20, zwłaszcza 15 [polski przekład: FRANKFURT 1997a), 21–39, zwłaszcza 32–33].

²⁵ Omówienie pojęcia utożsamienia zob. FRANKFURT 1976 [polski przekład: FRANKFURT 2001]; FRANKFURT 1987.

istnieje dopasowanie między wartościującą a motywacyjną preferencją wykonania tego działania²⁶.

I tym razem problem tkwi w tym, że teoria taka jest czysto strukturalna. Dopasowanie między składnikami różnych systemów preferencji można wywołać za pomocą elektronicznego oddziaływania, hipnozy, prania mózgu itd. Odpowiedzialność moralna jest zjawiskiem *historycznym* — zależy od rodzaju mechanizmu, który prowadzi do działania²⁷.

ZAKOŃCZENIE

Przedstawiłem szkic teorii mającej określać klasę działań, za które ludzie mogą racjonalnie podlegać moralnej pochwie lub naganie, nagrodzie lub karze. Twierdzę, że teoria ta ujmuje nasze jasne intuicyjne sądy na temat odpowiedzialności moralnej i pomaga uzgodnić ją z przyczynowym determinizmem. Z pewnością nie *udowodniłem*, że zachodzi zgodność między odpowiedzialnością a takim determinizmem. Moja strategia polegała raczej na wykazaniu, że przyjęte tu podejście pozwala kompatybilistom w kwestii moralnej odpowiedzialności i determinizmu uniknąć zobowiązań kompatybilisty w kwestii wolności postąpienia inaczej i determinizmu. Być może kompatybilista pierwszego rodzaju staje jeszcze przed innymi wyzwaniem, którym moje podejście, samo w sobie, nie potrafi sprostać.

Teoria, którą przedstawiłem, korzysta z ustaleń innych autorów i te ustalenia rozwija. Unika zarazem niektórych spośród najbardziej wyraźnych zarzutów formułowanych wobec teorii podobnego typu. Te zastrzeżenia mogą wydawać się przekonujące, jeśli nie „zachowa się w niezmienionej postaci” mechanizmu z faktycznej sekwencji albo przyjmie mocną, a nie słabą, zdolność reagowania na racje, albo też nie uwzględni się w teorii odpowiednich indeksów czasowych.

²⁶ Nie jestem pewien, czy sam Watson musi przyjąć wystarczalność dopasowania do odpowiedzialności moralnej. Z pewnością musi przyjąć, że sprawca jest wolny w takim stopniu, w jakim ma moc, by spowodować dopasowanie między wartościującymi i motywacyjnymi systemami. WATSON 1975, 216.

²⁷ Odpowiedzialność moralna przypomina pod tym względem takie pojęcia jak sprawiedliwość czy miłość do konkretnej osoby. W NOZICK 1974 [polski przekład: NOZICK 2010] autor argumentuje, że sprawiedliwość i miłość są pojęciami historycznymi raczej niż pojęciami z „bieżącego plastra czasu”. Czysto strukturalne podejścia do moralnej odpowiedzialności są nieadekwatne w sposób, który jest analogiczny do nieadekwatności podejść typu bieżącego plastra czasu do sprawiedliwości.

Chciałbym na koniec poczynić kilka sugestii na temat relacji między przedstawioną tu teorią moralnej odpowiedzialności a karą. Teoria odpowiedzialności moralnej musi wyjaśnić, czemu jedne stworzenia podlegają karze, a inne nie. Kara oznacza oczywiście, że ktoś zostaje potraktowany w nieprzyjemny sposób. Wpływa to na atrakcyjność podjęcia określonego działania. Inaczej mówiąc, kara oznacza taką reakcję wobec ludzi, na którą jest wrażliwy mechanizm kierujący ich działaniem. W moim przekonaniu kara jest odpowiednia jedynie dla stworzenia, które działa w oparciu o mechanizm dostrojony do bodźców, jakich ta kara dostarcza.

Nie chodzi o to, że uzasadnienie kary jest „konsekwencjalistyczne” — że kara zmienia zachowanie. (Ten rodzaj uzasadnienia sam w sobie z pewnością nie odróżnia kary od awersyjnego warunkowania). W istocie jest metafizycznie możliwe, by kompletny układ czyichś wyborów i działań na przestrzeni całego życia był „niezmienialny” ze względu na nieprzerwany ciąg sytuacji w stylu frankfurckim. (Jest nawet możliwe, że *żadne* ludzkie zachowanie nie może ulec zmianie, bo jest możliwe, że wszyscy ludzie podlegają kontrfaktycznym interwencjom typu frankfurckiego). Moje uzasadnienie jest niekonsekwencjalistyczne i „bezpośrednie”: kara jest odpowiednią reakcją na faktyczne działanie mechanizmów reagujących na racje. Uzasadniona kara obejmuje pewnego rodzaju „dopasowanie” między mechanizmem, który doprowadza do zachowania, a odpowiedzią na to zachowanie.

Teoria odpowiedzialności moralnej pozwala więc lepiej zrozumieć stosowność kary za pewne działania. Ale sama w sobie nie dostarcza pełnego opisu odpowiedniego *wymiaru* kary. I tak na przykład może się zdarzyć, że surowość kary za określone działanie jest mniejsza (albo większa) niż wielkość bodźca, na jaki reaguje mechanizm z faktycznej sekwencji. Jest to w pełni zgodne ze stwierdzeniem, że kara — będąc „źródłem racji” — jest we właściwy sposób skierowana do sprawców, którzy działają w oparciu o mechanizm wrażliwy na racje.

*Przełożyli Marcin Iwanicki
i Joanna Klara Teske*

Dr MARCIN IWANICKI — Katolicki Uniwersytet Lubelski Jana Pawła II, Wydział Filozofii, Katedra Historii Filozofii Nowożytnej i Współczesnej; adres do korespondencji: Al. Raławickie 14, 20-950 Lublin; e-mail: miwanick@gmail.com; ORCID: <https://orcid.org/0000-0002-3904-0291>.

Dr hab. JOANNA KLARA TESKE, prof. KUL — Katolicki Uniwersytet Lubelski Jana Pawła II, Wydział Nauk Humanistycznych, Instytut Literaturoznawstwa, Katedra Literatury i Kultury Angielskiej; adres do korespondencji: Al. Raławickie 14, Lublin 20-950; e-mail: jteske@kul.pl; ORCID: 0000-0002-3651-595X.

BIBLIOGRAFIA

- DAVIS, Lawrence H. 1979. *A Theory of Action*. Englewood Cliffs, N.J.: Prentice-Hall.
- DENNETT, Daniel C. 1978. „Mechanism of Responsibility”. W: Daniel C. DENNETT. *Brainstorms: Philosophical Essays on Mind and Psychology*, 233–255. Montgomery, Vt.: Bradford.
- DENNETT, Daniel C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: MIT Press.
- DRETSKE, Fred I. 1971. „Conclusive Reasons”. *Australasian Journal of Philosophy* 49: 1–22.
- FINGARETTE, Herbert. 1972. *The Meaning of Criminal Insanity*. Berkeley: University of California Press.
- FISCHER, John Martin. 1982. „Responsibility and Control”. *Journal of Philosophy* 79, nr 1: 24–40.
- FISCHER, John Martin. 1983. „Incompatibilism”. *Philosophical Studies* 43: 127–137.
- FISCHER, John Martin. 1985–1986. „Responsibility and Failure”. *Proceedings of the Aristotelian Society* 86: 251–272.
- FISCHER, John Martin. 1986a. „Van Inwagen on Free Will”. *Philosophical Quarterly* 36: 252–260.
- FISCHER, John Martin. 1986b. *My Way: Essays on Moral Responsibility*. Oxford: Cambridge University Press.
- FISCHER, John Martin. 1987. „Responsiveness and Moral Responsibility”. W: *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, red. Ferdinand Schoeman, 81–106. Cambridge: Cambridge University Press; przedruk w: FISCHER 1986b, 63–83.
- FISCHER, John Martin. 1988. „Freedom and Miracles”. *Nous* 22: 235–252.
- FRANKFURT, Harry. 1969. „Alternate Possibilities and Moral Responsibility”. *Journal of Philosophy* 46, nr 23: 829–839 [przekład polski: FRANKFURT 1997b].
- FRANKFURT, Harry. 1971. „Freedom of the Will and the Concept of a Person”. *Journal of Philosophy* 68: 5–20 [przekład polski: FRANKFURT 1997a].
- FRANKFURT, Harry. 1976. „Identification and Externality”. W: *The Identities of Persons*, red. Amélie Oksenberg Rorty, 239–251. Berkeley: University of California Press, 1976 [przekład polski: FRANKFURT 2001].
- FRANKFURT, Harry. 1987. „Identification and Wholeheartedness”. Rozdz. 2 w: *Responsibility, Character, and the Emotions*, red. Ferdinand Schoeman, 27–45. Cambridge: Cambridge University Press.
- FRANKFURT, Harry. 1997a. „Wolność woli i pojęcie osoby”. Przeł. Justyna Nowotniak. W: *Filozofia moralności. Postanowienie i odpowiedzialność moralna*, red. Jacek Hołówka, 21–39. Warszawa: Spacja-Aletheia].
- FRANKFURT, Harry. 1997b. „Alternatywne możliwości i odpowiedzialność moralna”. Przeł. Justyna Nowotniak. W: *Filozofia moralności. Postanowienie i odpowiedzialność moralna*, red. Jacek Hołówka, 163–174. Warszawa: Spacja-Aletheia.
- FRANKFURT, HARRY. 2001. „Identyfikacja i zewnętrzność”. Przeł. Sebastian Stecko. W: *Filozofia podmiotu*, red. Joanna Górnicka-Kalinowska, 363–375. Warszawa: Aletheia].
- GERT, Bernard, i Timothy J. DUGGAN. 1979. „Free Will as the Ability to Will”. *Nous* 13: 197–217.
- GINET, Carl. 1966. „Might We Have No Choice?”. W: *Freedom and Determinism*, red. Keith Lehrer, 87–104. New York: Random House.
- GLOVER, Jonathan. 1970. *Responsibility*. New York: Humanities Press, 1970.
- LAMB, James W. 1977. „On a Proof of Incompatibilism”. *Philosophical Review* 86: 20–35.
- LEVIN, Michael E. 1979. *Metaphysics and the Mind-Body Problem*. Oxford: Clarendon Press.
- LEWIS, David. 1981. „Are We Free to Break the Laws?”. *Theoria* 47: 113–121.

- LOCKE, John. 1690/1894. *Essay Concerning Human Understanding*. Project Gutenberg. Dostęp 15.11.2021. <https://www.gutenberg.org/ebooks/10615> (ks. I) i <https://www.gutenberg.org/ebooks/10616> (ks. II) [przekład polski: LOCKE 1955].
- LOCKE, John. 1955. *Rozważania dotyczące rozumu ludzkiego*. Przeł. Bolesław J. Gawecki. Warszawa: PWN.
- MACINTYRE, A[lasdair]. 1957. „Determinism”. *Mind* 66: 28–41.
- NEELY, Wright. 1974. „Freedom and Desire”. *Philosophical Review* 83: 32–54.
- NOZICK, Richard. 1974. *Anarchy, State, and Utopia*. New York: Basic Books [przekład polski: NOZICK 2010].
- NOZICK, Richard. 1981. *Philosophical Explanations*. Cambridge: Harvard University Press.
- NOZICK, Richard. 2010. *Anarchia, państwo i utopia*. Przeł. Michał Szczubiałka i Paweł Maciejko. Warszawa: Spacja-Aletheia].
- SHATZ, David. 1985. „Free Will and the Structure of Motivation”, *Midwest Studies in Philosophy* 10: 451-482.
- SHOPE, Robert K. 1984. „Cognitive Abilities, Conditionals, and Knowledge: A Response to Nozick”. *Journal of Philosophy* 81, nr 1: 29–48.
- SLOTE, Michael. 1982. „Selective Necessity and the Free-Will Problem”. *Journal of Philosophy* 82: 5–24.
- STRAWSON, Peter F. 1962. „Freedom and Resentment”. *Proceedings of the British Academy* 48: 1–25 [przekład polski: STRAWSON 2019].
- STRAWSON, Peter F. 2019. „Wolność a uraza”. *Przegląd Filozoficzny — Nowa Seria* 28: 35–57].
- VAN INWAGEN, Peter. 1975. „The Incompatibility of Free Will and Determinism”. *Philosophical Studies* 27: 185–199.
- VAN INWAGEN, Peter. 1978. „Ability and Responsibility”. *Philosophical Review* 87: 201–224; przedruk w: Peter VAN INWAGEN, *An Essay on Free Will*, 161–182. Oxford: Clarendon Press.
- VAN INWAGEN, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- WATSON, Gary. 1975. „Free Agency”. *Journal of Philosophy* 72: 205–220.
- WIGGINS, David. 1973. „Towards a Reasonable Libertarianism”. W: *Essays on Freedom of Action*, red. Ted Honderich, 31–61. Boston: Routledge and Kegan Paul.

ZDOLNOŚĆ REAGOWANIA NA RACJE A ODPOWIEDZIALNOŚĆ MORALNA

Streszczenie

Autor przedstawia model odpowiedzialności moralnej oparty na faktycznej sekwencji i pojęciu zdolności reagowania na racje, a następnie przeprowadza analogię między tym modelem a opracowanym przez Roberta Nozicka modelem wiedzy opartej na faktycznej sekwencji, oraz wprowadza pojęcie semikompatybilizmu.

Słowa kluczowe: wolna wola; odpowiedzialność moralna; zdolność reagowania na racje.

RESPONSIVENESS AND MORAL RESPONSIBILITY

Summary

The author presents a model of moral responsibility based on the actual sequence and the notion of reason-responsiveness, and draws an analogy between this model and Robert Nozick's model of knowledge based on the actual sequence. In addition, the concept of semicompatibilism is introduced and explained.

Keywords: free will; moral responsibility; reason-responsiveness.

Information about the Author: Prof. JOHN MARTIN FISCHER — Professor of Philosophy at the University of California, Riverside; correspondence address: Department of Philosophy, 900 University Ave., Riverside, CA 92521; e-mail: john.fischer@ucr.edu; ORCID: <https://orcid.org/0000-0002-1685-4743>.

Information about the Translators:

MARCIN IWANICKI, PhD — The John Paul II Catholic University of Lublin, Faculty of Philosophy, Department of the History of Modern and Contemporary Philosophy; correspondence address: Al. Raławickie 14, 20-950 Lublin; e-mail: miwanick@gmail.com; ORCID: <https://orcid.org/0000-0002-3904-0291>.

Dr. habil. JOANNA KLARA TESKE, Associate Professor at KUL — The John Paul II Catholic University of Lublin, Faculty of Humanities, Institute of Literary Studies, Department of English Literature and Culture; correspondence address: Al. Raławickie 14, Lublin 20-950; e-mail: jteske@kul.pl; ORCID: <https://orcid.org/0000-0002-3651-595X>.