

A d r i a n K u ź n i a r

Kompatybilizm a postawa obiektywna w świecie mechanicznych pomarańczy

Słowa kluczowe: *kompatybilizm, postawa obiektywna, moralne uczucia reaktywne, odpowiedzialność moralna, ekspresywizm*

Niniejszemu artykułowi towarzyszą dwa podstawowe cele badawcze. Pierwszym jest rekonstrukcja i krytyka wylaniającej się z tekstu Petera Strawsona pt. *Freedom and Resentment* (1962)¹ próby sprowadzenia stanowiska antykompatybilistycznego do absurdu. Drugi z celów badawczych polega na zarysowaniu pewnej wersji kompatybilizmu – wersji zarówno unikającej zarzutów Strawsona wobec tradycyjnych teorii tego typu, jak i umożliwiającej głębszą analizę różnych możliwych interpretacji sporu kompatybilistów z ich przeciwnikami.

W punkcie 1 artykułu dokonuję wzmiankowanej rekonstrukcji argumentu *ad absurdum* przeciwko Strawsonowskiemu „pesymiście”. W punkcie 2 zgłaszam wątpliwości dotyczące prawdziwości jednej z przesłanek zrekonstruowanego argumentu, tj. twierdzenia o niemożliwości powszechnego zajęcia postawy obiektywnej wobec wszystkich jednostek wykraczających przeciwko porządkowi moralnemu. Swoje wątpliwości popieram m.in. eksperymentem myślowym nawiązującym do fikcyjnej techniki Ludovycka z powieści *Mechaniczna pomarańcza* Anthony’ego Burgessa (1962)². W kolejnym punkcie zarysowuję hipotezę funkcjonalistyczną, udzielającą wsparcia mojemu przewidywaniu dotyczącemu

Adrian Kuźniar, Uniwersytet Warszawski, Instytut Filozofii, ul. Krakowskie Przedmieście 3, 00-927 Warszawa; e-mail: adrian_kuzniar@uw.edu.pl, ORCID: 0000-0002-0428-8203.

¹ Dalej cytuję wydanie: Strawson 1993.

² Dalej cytuję wydanie: Burgess 2017.

ewolucji postaw w świecie mechanicznych pomarańczy, a przez to podbudowującą przedstawioną krytykę Strawsonowskiej *reductio ad absurdum*. Hipoteza ta nie dostarcza jednak racji na rzecz antykompatybilizmu; przeciwnie – ma charakter kompatybilistyczny. W punkcie 4 przedstawiam ekspresywistyczne analizy pojęć kluczowych dla sądów istotnych dla praktyki przypisywania odpowiedzialności moralnej. Analizy te odwołują się do planów dla moralnych uczuć reaktywnych, co harmonizuje z wywodami Strawsona. W następnym punkcie artykułu omawiam i odpowiadam na zarzuty, jakie Strawson stawia tradycyjnemu podejściu kompatybilistycznemu. Wykazuję, że nie są one słuszne w stosunku do tej wersji kompatybilizmu, którą tworzą proponowane analizy pojęciowe wraz z hipotezą funkcjonalistyczną. W ostatnim punkcie przedstawiam różne możliwe interpretacje sporu o kompatybilizm oraz rozważam relacje, w jakich broniony przeze mnie w tym artykule pogląd pozostaje do tychże interpretacji.

1. *Reductio ad absurdum* przeciwko pesymizm

W artykule Strawsona *Freedom and Resentment* znajdujemy kilka argumentów przeciwko antykompatybilizm, określanego tam jako „pesymista”. W bieżącym punkcie chciałbym zrekonstruować pewną linię rozumowania wymierzonego w pesymistę, której ślady możemy w mojej opinii wytropić w tekście Strawsona. Jest to argument *reductio ad absurdum* i jako taki polega na sprowadzeniu twierdzenia pesymisty do niedorzeczności. Pierwszą jego przesłanką jest jedno z definiujących antykompatybilizm twierdzeń, zgodnie z którym determinizm wyklucza odpowiedzialność moralną:

(1)

Jeśli nasz świat jest deterministyczny, to żaden podmiot wykraczający w naszym świecie przeciwko porządkowi moralnemu nie jest podmiotem moralnie odpowiedzialnym.

Drugą przesłanką w rekonstruowanym rozumowaniu jest substancjalne i charakterystyczne dla poglądów Strawsona twierdzenie z dziedziny etyki normatywnej:

(2)

Jeśli podmiot wykraczający przeciwko porządkowi moralnemu nie jest podmiotem moralnie odpowiedzialnym, to praktycznie racjonalne jest zajęcie wobec niego postawy obiektywnej.

Postawa obiektywna wobec innego to w interpretacji Strawsona postawa, w ramach której traktujemy innego jako przedmiot polityki społecznej lub

terapii, jako kogoś, kogo można i w określonych okolicznościach należy kontrolować lub nim zarządzać³. Zajęcie wobec innego postawy obiektywnej oznacza zawieszenie lub zniesienie w stosunku do niego wszystkich postaw i uczuć reaktywnych. Inny jest właściwym obiektem tych ostatnich tylko wówczas, gdy przyjmujemy wobec niego postawę uczestniczącą. Wtedy może być obiektem naszej wdzięczności, urazy, przebaczenia, moralnego oburzenia, gniewu itp.

Z przesłanek (1) i (2) wynika następujący wniosek:

(3)

Jeśli nasz świat jest deterministyczny, to praktycznie racjonalne jest zajęcie postawy obiektywnej wobec wszystkich podmiotów wykraczających w naszym świecie przeciwko porządkowi moralnemu.

Strawson przyjmuje, że:

(4)

Praktyczna racjonalność implikuje praktyczną możliwość⁴.

Na podstawie (3) i (4) otrzymujemy:

(5)

Jeśli nasz świat jest deterministyczny, to przyjęcie postawy obiektywnej wobec wszystkich podmiotów wykraczających w naszym świecie przeciwko porządkowi moralnemu jest praktycznie możliwe.

Twierdzenie (5) stanowi poszukiwaną niedorzeczność; Strawson nie ma bowiem wątpliwości, że powszechne przyjęcie postawy obiektywnej – przyjęcie jej wobec wszystkich podmiotów działających – jest praktycznie niemożliwe, bez względu na to, czy nasz świat aktualny jest światem deterministycznym, czy wręcz przeciwnie – indeterministycznym. Przekonanie to ujmijmy pod postacią **twierdzenia o niemożliwości**:

(6)

Bez względu na to, czy nasz świat jest deterministyczny, czy indeterministyczny, jest nieprawdą, że przyjęcie postawy obiektywnej wobec wszystkich podmiotów wykraczających w naszym świecie przeciwko porządkowi moralnemu jest praktycznie możliwe.

³ Strawson 1993, s. 52.

⁴ Strawson 1993, s. 54, 55, 60. W kontekście artykułu Strawsona przesłankę (4) możemy uznać za po prostu alternatywną wersję stylistyczną twierdzenia głoszącego, że powinność zakłada możliwość.

W dyskutowanym artykule Strawson bardzo często kładzie nacisk na nieco różne od (6) twierdzenie, argumentując, że praktycznie niemożliwe jest przyjęcie postawy obiektywnej w stosunku do wszystkich podmiotów **wskutek** akceptacji twierdzenia o deterministycznym charakterze świata aktualnego. W tej wersji twierdzenie o niemożliwości nie implikuje twierdzenia (6): z tego, że przyjęcie postawy obiektywnej nie jest możliwe na skutek powzięcia pewnego konkretnego przekonania, nie wynika, że jest to niemożliwe bez względu na okoliczności. Pomimo to w swojej rekonstrukcji przyjmuję (6); czynię tak z dwóch zasadniczych względów. Po pierwsze, ostrożniejsze sformułowanie twierdzenia o niemożliwości nie stoi w sprzeczności z (5), co oznaczałoby, że argument przeciwko pesymiście zawodziłby już od strony formalnej. Po drugie, enuncjacje Strawsona jednoznacznie sugerują, że trwale zajęcie postawy obiektywnej wobec wszystkich uważa on za niemożliwe w każdych warunkach, **nawet wtedy**, gdybyśmy akceptowali twierdzenie stanowiące *prima facie* rację na rzecz takiego kroku⁵.

W związku z akceptacją (6) należy odrzucić jedną z przesłanek prowadzących do twierdzenia (5), tj. przesłankę (1), (2) lub (4). Na gruncie proponowanej rekonstrukcji argumentu Strawsona (6) falsyfikuje twierdzenie pesymisty, zatem ostatecznym wnioskiem z rozumowania jest negacja (1):

(7)

Nieprawda, że jeśli nasz świat jest deterministyczny, to żaden podmiot wykraczający w naszym świecie przeciwko porządkowi moralnemu nie jest podmiotem moralnie odpowiedzialnym.

2. Wątpliwości wokół twierdzenia o niemożliwości

Twierdzenie o niemożliwości uznaję za wątpliwe. Swoją argumentację zmierzającą do podważenia go rozpocznę od eksperymentu myślowego, którego inspiracją jest fikcyjna technika Ludovycka z *Mechanicznej pomarańczy* Anthony'ego Burgessa⁶. Główny bohater tej powieści, Alex, traci możliwość realizacji swoich zbrodniczych pragnień wskutek przejścia nowatorskiej terapii. W rezultacie tej terapii doświadcza nieznośnego fizycznego bólu, mdłości i poczucia, że zaraz umrze, ilekroć pojawiają się w nim naganne pod względem moralnym pragnienia i ilekroć próbuje pragnienia te przekuć w czyn. Mimowolnie Alex staje się przyzwoitym, a nawet więcej niż przyzwoitym

⁵ Strawson 1993, s. 54, 55, 60.

⁶ Głośną ekranizacją dzieła Burgessa jest film z 1971 r. pod tym samym tytułem w reżyserii Stanleya Kubricka. W filmie tym skutki terapii Ludovycka zostały ukazane bardzo dobitnie.

człowiekiem: w obawie przed cierpieniem stara się eliminować złe pragnienia i robi wszystko, by zachowywać się moralnie i altruistycznie. Powieściowy doktor Brodzki tak charakteryzuje jego stan:

Nasz obiekt, jak państwo widzicie, ulega przymusowi dobra przez to, paradoksalnie, że ulega przymusowi zła. Intencja zadania gwałtu natychmiast łączy się z przemożnymi doznaniem fizycznego dyskomfortu. Ażeby mu zapobiec, obiekt musi błyskawicznie przestawić swój odruch na jego diametralne przeciwieństwo⁷.

Nieco dalej:

– Będzie z niego dobry chrześcijanin! – wykrzyknął doktor Brodzki. – Właśnie taki! gotów nadstawić drugiego policzka! raczej da się ukrzyżować, niżby sam kogoś ukrzyżował, i do głębi serca się wstrząśnie na myśl choćby o zabiciu muchy⁸.

Alex określa to, kim stał się po przejściu terapii, mianem „mechanicznej pomarańczy”⁹.

Wyobraźmy sobie, że terapia Ludovicka przestaje być fikcją i staje się dostępną. Wyobrażenie to nie wydaje się nazbyt ekstrawaganckie w dobie zachodzącego, imponującego postępu technicznego. Załóżmy także, że w rezultacie demokratycznej decyzji wszyscy jesteśmy przewencyjnie, tj. zanim jeszcze uczynimy cokolwiek złego, poddawani owej terapii. Niestety okazuje się, że we wdrożonym zautomatyzowanym systemie zdarzają się błędy: nielicznych terapia omija. W konsekwencji pewne jednostki nadal dopuszczają się transgresji moralnych, przy czym czynią zło na bazie tych samych mechanizmów, które stały za wykroczeniami moralnymi i zbrodniami w dotychczasowej historii świata. Pytaniem zasadniczym w kontekście rozważanego scenariusza jest pytanie o to, jak w zarysowanych okolicznościach reagowalibyśmy na tych nielicznych przestępców – czy przypisywalibyśmy im odpowiedzialność moralną i przejawialibyśmy wobec nich postawy reaktywne w rodzaju moralnego oburzenia, urazy itp.? W moim przekonaniu odpowiedź na to pytanie jest

⁷ Burgess 2017, s. 139.

⁸ Burgess 2017, s. 141–142.

⁹ Oryginalny tytuł książki Burgessa to *A Clockwork Orange*. Wyrażenie *a clockwork orange* należałoby przełożyć raczej jako „nakręcana pomarańcza”, nie zaś „mechaniczna”. Polski tłumacz książki, Robert Stiller, był w pełni tego świadomy. Wyjaśnia, że zdecydował się na przekład z przymiotnikiem „mechaniczna” ze względów handlowych, a następnie po to, by odróżnić od siebie dwie wersje przekładu: tę, której podstawą był futurologiczny zrusyfikowany język polski (w niej występuje przymiotnik „mechaniczna”), od wersji opartej na futurologicznej zanglicyzowanej polszczyźnie (w tej ostatniej wersji dzieło Burgessa nosi tytuł *Nakręcana pomarańcza*). Zob. Stiller 2017, s. 211–212. W swoim artykule pozostają przy bardziej rozpowszechnionym tłumaczeniu.

przecząca, a moje **przewidywanie dotyczące świata mechanicznych pomarańczy** jest następujące: z czasem bylibyśmy skłonni do zawieszenia wobec takich jednostek moralnych uczuć i postaw reaktywnych, a nasza postawa wobec nich w kontekście ich transgresji stawałaby się postawą obiektywną – sądzilibyśmy, że trzeba uszczelnić i doskonalić system terapii, a nie reagować moralnym oburzeniem i urazą na nieliczne podmioty dopuszczające się złych moralnie działań.

Zło moralne nie polega jedynie na niewłaściwym działaniu; może polegać także na zawinionym zaniechaniu, wynikającym np. z zawinionej niewiedzy. W kategoriach moralnych możemy myśleć także o stanach emocjonalnych lub braku pewnych uczuć. Możemy na przykład sądzić, że brak odczuwania żalu w sytuacji bycia przypadkowym świadkiem śmiertelnego wypadku komunikacyjnego jest moralnie niewłaściwy. Świat mechanicznych pomarańczy powinniśmy zatem wyobrazić sobie jako taki, w którym przejście przez terapię gwarantuje nieobecność w podmiocie i jego działaniu wszystkiego, co zasługuje na negatywną ocenę moralną. Elementem takiej fikcyjnej terapii mogłaby być m.in. edukacja połączona z farmakologicznym uwrażliwieniem i polepszeniem zdolności koncentracji, czego rezultatem byłaby eliminacja zjawiska zawinionej niewiedzy. Sednem proponowanego eksperymentu myślowego jest pytanie o ewolucję naszych postaw wobec tych, którzy w świecie mechanicznych pomarańczy nie zostali z jakichś względów poddani terapii i w dodatku uczynili coś moralnie złego. W moim przekonaniu w opisywanych warunkach postawę uczestniczącą wobec takich jednostek zastąpiłaby z czasem postawa w pełni obiektywna.

Sądzę, że istnieją pewne racje empiryczne przemawiające na rzecz sformułowanego przewidywania. Jeśli przyjrzymy się w perspektywie diachronicznej polityce karnej i filozofii kary implementowanych w praktyce w zamożnych krajach Europy Zachodniej, to dostrzeżemy wyraźne osłabienie retributywistycznych aspektów kary, wzmocnienie zaś jej aspektów konsekwencjalistycznych. Kara jest coraz rzadziej postrzegana w kategoriach sprawiedliwej i należytej odpłaty, częściej natomiast niż w przeszłości myśli się o niej jako o narzędziu resocjalizacji i środka zabezpieczającym społeczeństwo przed niebezpiecznymi jednostkami. Nie przesądzając o sile wzmiankowanej tendencji, jej manifestacje możemy zaobserwować m.in. w zniesieniu kary śmierci, zapewnianiu humanitarnych warunków odbywania kary, rezygnacji z kary na rzecz izolacji i terapii w zakładzie leczniczym.

Kolejnym świadectwem na korzyść mojego przewidywania jest dokonujący się w krajach Zachodu proces obiektywizacji postawy wobec ludzi uzależnionych – zwłaszcza narkomanów i alkoholików. Obiektywizacja ta rozciąga się także na sytuacje, w których uzależnienie jest przyczyną czynów nagannych moralnie. Uzależnienia coraz częściej traktowane są jako choroby wymagające

leczenia, a nie – jak dotychczas – jako przejawy zepsutego moralnie charakteru. Ogólnie możemy stwierdzić, że w rozwiniętych krajach Zachodu obserwujemy wzmocnienie udziału postawy obiektywnej wobec osób wykraczających przeciwko porządkowi normatywnemu.

Przyjmijmy, że przedstawiony eksperyment myślowy oraz przytoczone świadectwa dają podstawę do podania w wątpliwość prawdziwości (6), a więc prawdziwości Strawsonowskiego twierdzenia o niemożliwości. Załóżmy, że nasza postawa wobec podmiotów dopuszczających się czynów niemoralnych będzie ewoluować w kierunku jej uniwersalnej obiektywizacji, a wraz z nią zanikać będą inne postawy i uczucia reaktywne. Wszystko to nie oznacza, rzecz jasna, przeobrażenia ludzi w niedostępnych afektom stoickich mędrców; proces ten pozostawia miejsce na bogate życie emocjonalne i szeroki wachlarz stanów afektywnych. Strawson wprost to przyznaje¹⁰. W tym kontekście ciekawym zagadnieniem pozostaje los takich uczuć, jak np. żal sprawcy¹¹. Załóżmy, że doznawanie tego uczucia we właściwych okolicznościach nie jest wymaganiem moralnym (jeśli ma charakter moralny, wówczas na mocy opisu świata mechanicznych pomarańczy wszyscy poddani terapii odczuwają ten stan w stosownych okolicznościach). Czy w tej sytuacji w naszym wyobrażonym świecie powinniśmy oczekiwać zaniku także tego uczucia? To, że obiektywizacja postawy nie będzie miała żadnego wpływu na występowanie żalu świadka – tj. żalu, że zaszedł taki, a nie inny stan rzeczy – jest jasne. Williamsowski żal sprawcy to jednak coś więcej niż żal świadka. Z jednej strony jest on właściwy w sytuacjach, w których podmiot ze względu np. na pechowy zbieg okoliczności **nie** ponosi odpowiedzialności moralnej za konkretne swoje działanie (a więc nie ma podstaw, by czuł się winny z powodu tego, co się stało); z drugiej jednak strony nie oczekujemy, by ten typ żalu odczuwały osoby, które uważamy za takie, którym z zasadniczych względów nie powinniśmy przypisywać odpowiedzialności. Chociaż świat mechanicznych pomarańczy to świat, w którym przestajemy postrzegać działających w kategoriach podmiotów odpowiedzialnych, to nie sądzę, by opis tego świata dawał podstawy do oczekiwania, że żal sprawcy będzie zanikał. Sytuacje, w których obecnie pojawia się w nas to uczucie i jego ukierunkowane na innych oczekiwanie, nadal będą się zdarzać, natomiast zajęcie postawy obiektywnej wobec bliźniego nie stoi w żadnym napięciu z uczynieniem dyspozycji do odczuwania żalu świadka podstawą selekcji kręgu przyjaciół, współpracowników, partnerów itp. Możemy po prostu chcieć otaczać się ludźmi zdolnymi do tego uczucia i zarazem nie traktować tej zdolności jako zalety moralnej. Stopień, w którym obecnie posiadamy tę preferencję, pozostałby niezmienny.

¹⁰ Strawson 1993, s. 52.

¹¹ Por. Williams 1999.

Preferencja ta byłaby czynnikiem podtrzymującym występowanie uczucia żalu sprawy w świecie mechanicznych pomarańczy.

Czy upadek zrekonstruowanego argumentu *reductio ad absurdum* przeciwko pesymiście – upadek będący rezultatem zanegowania przesłanki (6) – należy potraktować jako wzmocnienie stanowiska antykompatybilistycznego? W moim przekonaniu odpowiedź na to pytanie jest – co może wydawać się zaskakujące – negatywna.

3. Kompatybilistyczna hipoteza funkcjonalistyczna

Utrzymuję, że najlepszym wyjaśnieniem ewentualnej trafności przewidywania dotyczącego świata mechanicznych pomarańczy jest tradycyjna kompatybilistyczna hipoteza funkcjonalistyczna (odtąd w skrócie: hipoteza funkcjonalistyczna) i to ona najlepiej wyjaśnia tendencje, które dostarczają empirycznego wsparcia wspomnianemu przewidywaniu. Według tej hipotezy:

(8)

Funkcją postawy uczestniczącej przejawiającej się w praktyce przypisywania odpowiedzialności moralnej i związanej z nią dynamice moralnych uczuć i postaw reaktywnych jest minimalizowanie wykroczeń przeciwko porządkowi normatywnemu w grupie społecznej.

Racjonalność akceptacji hipotezy funkcjonalistycznej nie jest jednak uzależniona od trafności przewidywania dotyczącego świata mechanicznych pomarańczy. Jest tak dlatego, że najprawdopodobniej postawa uczestnicząca, manifestująca się w postawach i uczuciach reaktywnych, jest biologiczną adaptacją naszego gatunku. Jest ona adaptacyjnym mechanizmem psychologicznym realizującym określoną funkcję ewolucyjną – funkcję opisywaną przez hipotezę funkcjonalistyczną. Zajmowanie i zawieszanie tej postawy wobec innych jest uzależnione od takich własności sytuacji, które w ewolucyjnej historii naszego gatunku były skorelowane z efektywnością (*resp.* nieskutecznością) postawy uczestniczącej w realizacji jej funkcji. W kontekście własności zawieszających postawę uczestniczącą były to rozmaite własności sfery wolicjonalnej, psychicznej i poznawczej podmiotu. Wśród tych własności nie było jednak ogólnej własności w rodzaju: bycie podmiotem, wobec którego zajęcie postawy uczestniczącej jest nieoptymalnym instrumentem minimalizowania jego transgresji. Mechanizmy ewolucyjne są najczęściej pośrednie i takim mechanizmem jest również postawa uczestnicząca – jej inhibicja jest aktualizowana w rezultacie stwierdzenia egzemplifikacji własności w zwykłych okolicznościach skorelowanych z tą własnością, która jest kluczowa z punktu widzenia funkcji ewo-

lucyjnej rozpatrywanego mechanizmu. Jednak stwierdzenie obecności samej tej kluczowej własności nie wywołuje zawieszenia postawy – dyskutowany mechanizm nie jest bezpośrednio wrażliwy na jej egzemplifikacje.

Opisane powyżej okoliczności sprawiają, że możemy wyobrazić sobie sytuacje, w których postawa uczestnicząca nie będzie zawieszana, mimo że nie będzie realizować swojej funkcji ewolucyjnej. Może tak być wówczas, gdy będziemy mieli do czynienia z sytuacjami, które nie będą posiadały wspomnianych własności pośredniczących, tj. tych, które w historii ewolucyjnej naszego gatunku warunkowały inhibicję postawy uczestniczącej, służąc za wystarczająco wiarygodne wskaźniki efektywności lub nieskuteczności postawy uczestniczącej. Tego rodzaju sytuacją jest właśnie świat mechanicznych pomarańczy, gdzie nieliczni przestępcy to w zdecydowanej większości ludzie całkowicie „normalni” – w pełni władz wolicjonalnych, psychicznych i poznawczych. Na mocy omówionych powyżej racji wystąpienie takiego zjawiska nie będzie jednak falsyfikatorem dla hipotezy funkcjonalistycznej.

Pomimo zgłoszonego zastrzeżenia sądzę, że sformułowane przewidywanie dotyczące świata mechanicznych pomarańczy jest z większym prawdopodobieństwem prawdziwe niż fałszywe. Należy spodziewać się, że w świecie tym, w obliczu widocznych dla każdego sukcesów terapii Ludovycka, będzie rozpowszechniać się wiedza dotycząca neurofizjologicznych przyczyn zachowania i ludzkich działań. Tymczasem, jak pokazują badania, ludzie mają skłonność do traktowania wyjaśnień neurofizjologicznych jako wykluczających wyjaśnienia psychologiczne¹². W konsekwencji często wykazują oni tendencję do klasyfikowania zachowań wyjaśnianych w kategoriach neurofizjologicznych jako niebędących w rzeczywistości dziełem podmiotu – niebędących skutkiem jego autentycznych pragnień, intencji i planów. To z kolei sprawia, że zaliczają tego typu przypadki do grupy tych warunków, w których podmiot nie jest odpowiedzialny moralnie za to, co się stało. Po drugie, co znacznie bardziej spekulatywne – przypuszczam, że refleksja nad racjonalnością praktyki przypisywania odpowiedzialności moralnej i karania prowadzona w świecie mechanicznych pomarańczy będzie faworyzować rozwiązania zgodne z hipotezą funkcjonalistyczną, promujące zajmowanie postawy obiektywnej. Sądzę, że będzie tak nawet wtedy, gdy zakorzenione w nas ewolucyjnie heurystyki psychologiczne będą działać w przeciwnym kierunku, tj. w kierunku traktowania przestępców, których ominęła terapia Ludovycka, jako w pełni odpowiedzialnych moralnie za swoje czyny. Nie jesteśmy bezradni w obliczu implementowanych w nas w rezultacie doboru naturalnego mechanizmów psychologicznych; np. wielu z nas ogranicza w swojej diecie cukier i tłuszcz, skutecznie walczy z arachnofobią, stara się anonimowo pomagać ludziom w odległych krajach itd. Niekiedy

¹² Por. Nadelhoffer i in. 2013, s. 198–199.

pojawiają się po prostu względy, które biorą górę nad tym, za czym przemawiają pragnienia należące do wąsko wyspecjalizowanych ewolucyjnych modułów psychologicznych. Myślę, że wzgląd na ekonomię systemu kar i efektywność ekspresji moralnych uczuć reaktywnych skłoniłby nas w świecie mechanicznych pomarańczy do zajmowania wobec wszystkich przestępców – w związku z ich przestępczymi działaniami – postawy obiektywnej.

Zatem ostatecznie, na gruncie hipotezy funkcjonalistycznej należy oczekiwać, że:

- (A) Znaczenie instrumentu, jakim jest postawa uczestnicząca, zmniejsza się i będzie zmniejszać się tam, gdzie postawa ta jest szczególnie nieskuteczna w redukcji liczby transgresji, oraz tam, gdzie pojawiają się skuteczniejsze środki utrzymywania porządku normatywnego.
- (B) Wymienione w (A) obszary będą obszarami ekspansji postawy obiektywnej.
- (C) Opisowany proces będzie przebiegał z licznymi i poważnymi zakłóceniami z uwagi na trwałość mechanizmów psychologicznych o genezie adaptacyjnej.

Akceptacja hipotezy funkcjonalistycznej może sugerować, że powracamy do tego błędu kompatybilistów, który piętnuje Strawson, gdy krytykuje optymistę. Tak nie jest. Zanim jednak będę mógł to wykazać i odeprzeć konkretne zarzuty Strawsona, muszę dysponować eksplikacjami podstawowych pojęć zaangażowanych w praktykę przypisywania odpowiedzialności moralnej.

4. Analizy pojęciowe

Proponowane przeze mnie analizy pojęciowe są analizami ekspresywistycznymi nawiązującymi do teorii metaetycznej Allana Gibbarda¹³. Oznacza to po pierwsze, że kluczowe wyrażenia moralne rozjaśniane są za pomocą opisu stanów umysłu, w których znajduje się osoba dokonująca asercji sądu o treści charakteryzowanej przy użyciu wyjaśnianych wyrażen. Wyjaśnienie znaczenia terminów nie polega tutaj na przytoczeniu ich semantycznych ekwiwalentów. Po drugie, przynajmniej jeden stan umysłu opisywany w eksplanansie nie jest przekonaniem w wąskim znaczeniu, tj. stanem umysłu o kierunku odpowiedniości od umysłu do świata; jest natomiast stanem akceptacji pewnego planu, tj. stanu umysłu o kierunku odpowiedniości od świata do umysłu¹⁴. Po trzecie, plany, o których mowa, są planami dla moralnych uczuć reaktywnych, co bar-

¹³ Gibbard 1990.

¹⁴ Searle 1985, s. 7–9.

dzo dobrze harmonizuje z treścią poglądów Strawsona. Zawarte w bieżącym punkcie objaśnienia do zaprezentowanych poniżej eksplikacji (9)–(11) mają jedynie wstępny charakter. Dalsze i pogłębione zrozumienie ich treści stanie się możliwe w toku dyskusji, którą podejmę w punkcie 5 niniejszego artykułu. Tam postaram się odeprzeć główne zarzuty formułowane przez Strawsona pod adresem klasycznego stanowiska kompatybilistycznego, tzn. postaram się wykazać, że moja wersja tego poglądu nie jest na te zarzuty narażona.

Rozpocznijmy od wyjaśnienia natury stanu umysłu polegającego na sądzie- niu, że pewien typ czynów jest moralnie zły.

(9)

Osoba *S* sądzi, że typ *T* czynów jest **moralnie zły** ztw., gdy:

(A) *S* akceptuje plan, by mieć poczucie winy w sytuacji, w której czyni *T* &

(B) *S* akceptuje plan, by odczuwać moralne oburzenie wobec każdego, kto dopuszcza się *T*,

przy czym wzmiankowane plany zawierają wyjątek:

(C) z wyjątkiem sytuacji, w której działający znajduje się w okoliczno- ściach znoszących odpowiedzialność moralną.

Zatem wyjaśniany stan umysłu jest stanem akceptacji planów dla uczuć moral- nych – poczucia winy i moralnego oburzenia – planów na okoliczność zaistnienia czynu określonego typu. Plany te zawierają klauzulę, zgodnie z którą nie pla- nuje się odczuwania tych uczuć, jeśli czyn należący do wzmiankowanego typu został dokonany przez podmiot w okolicznościach znoszących odpowiedzialność moralną. Zauważmy, że zgodnie z zaproponowaną analizą sądenie, że coś jest moralnie złe, nie polega eksplikacji w terminach przypisania przedmiotowi pewnej własności. Jest to charakterystyczna cecha ujęć ekspresywistycznych.

Analiza (9) jest niepełna, dopóki nie wyjaśnimy natury stanu umysłu będą- ącego sądem w sensie psychologicznym, że pewna okoliczność należy do zbioru okoliczności znoszących odpowiedzialność moralną za czyn danego typu:

(10)

Osoba *S* sądzi, że *O* należy do zbioru *M* okoliczności **znoszących odpo- wiedzialność moralną** za czyn typu *T* ztw., gdy:

(A) *S* akceptuje plan, by nie odczuwać poczucia winy za *T* w *O* &

(B) *S* akceptuje plan, by nie odczuwać moralnego oburzenia wobec nikogo, kto w *O* dopuszcza się *T* &

(C) *S* akceptuje plan, by odczuwać poczucie winy i moralne oburzenie za *T* w sytuacji, w której do *T* dochodzi w okolicznościach innych niż należące do *M*.

Według (10) eksplikowany stan umysłu, podobnie jak sążenie, że coś jest moralnie złe, polega na akceptacji planów dla uczuć moralnych w związku z działaniem typu T . W tym przypadku są to plany, by uczuć tych nie odczuwać, jeśli podmiot działał w sytuacji określonego rodzaju. Akceptacji tych planów towarzyszy akceptacja jeszcze jednego planu, a mianowicie, by odczuwać wzmiankowane uczucia we wszystkich tych warunkach, które nie należą do wzmiankowanego rodzaju, lecz w których mamy do czynienia z działaniem egzemplifikującym typ T . Uwzględnienie akceptacji tego ostatniego planu zdaje sprawę z faktu, że sądy na temat okoliczności znoszących odpowiedzialność za czyn egzemplifikujący typ T są sądami zakładającymi, że czyny należące do tego typu są moralnie złe.

Obecnie posiadamy już wszystko, co niezbędne, do wyjaśnienia sądów w sensie psychologicznym, których przedmiotem jest odpowiedzialność moralna za działanie:

(11)

Osoba S_1 sądzi, że S_2 jest **odpowiedzialna moralnie** za czyn typu T ztw., gdy S_1 sądzi, że:

(A) czyny T są moralnie złe &

(B) S_2 dopuściła się czynu należącego do T &

(C) dla żadnej okoliczności O należące do zbioru M okoliczności znoszących odpowiedzialność moralną za czyn typu T nie jest prawdą, że S_2 znajdowała się w O .

To, na czym polega sążenie, że czyny T są moralnie złe, wyjaśniłem w (9). Stan umysłu scharakteryzowany w punkcie (B) powyżej jest sądem, w którym dokonujemy atrybucji sprawstwa. Z kolei stan umysłu z punktu (C) możemy zrozumieć dzięki eksplikacji (10).

5. Odpowiedź na zarzuty

W bieżącym punkcie postaram się zrekonstruować formułowane przez Strawsona zarzuty wobec tradycyjnej wersji teorii kompatybilistycznej i wykazać, że zarzuty te nie są trafne w zastosowaniu do stanowiska będącego koniunkcją przedstawionych analiz pojęciowych i hipotezy funkcjonalistycznej.

5.1. Pierwszy zarzut

Pierwszym zarzutem Strawsona przeciwko optymizmowi jest zwrócenie uwagi na fakt, że pojęcia centralne dla praktyki przypisywania odpowiedzialności moralnej nie podlegają eksplikacji za pomocą takich pojęć, jak pojęcie sku-

teczności sankcji, kontroli, terapii itp., lecz za pomocą pojęć moralnych uczuć i postaw reaktywnych¹⁵. Przedstawione przeze mnie analizy pojęciowe w sposób oczywisty tego zarzutu unikają: głównymi pojęciami zaangażowanymi w tę analizę są pojęcia moralnych uczuć reaktywnych – moralnego oburzenia i poczucia winy. Sądy, których przedmiotem jest odpowiedzialność moralna, są w ramach zaprezentowanego ujęcia stanami akceptacji planów dla wspomnianych uczuć reaktywnych.

5.2. Drugi zarzut

Według Strawsona, z teorii optymisty w jej tradycyjnej wersji zdaje się wynikać, że powinniśmy przyjmować postawę obiektywną wobec każdego i że uzasadniony jest jedynie taki wzorzec atrybucji odpowiedzialności, który jest skutecznym narzędziem podtrzymywania porządku normatywnego¹⁶.

Odpowiadając na ten zarzut, zaznaczmy, że zarysowana przeze mnie składowa teorii kompatybilistycznej pod postacią hipotezy funkcjonalistycznej jest teorią czysto opisową, dotyczącą funkcji istnienia i kształtu praktyki przypisywania odpowiedzialności. Funkcja ta jest rozumiana w perspektywie genealogicznej, nie zaś normatywnej. Wskazuję więc na funkcję, której realizowanie wyjaśnia obserwowaną praktykę. Hipoteza funkcjonalistyczna nie jest jednak teorią normatywną, która dotyczyłaby racji i uzasadnienia sądów o odpowiedzialności. Imputowane tradycyjnemu kompatybilistycie przez Strawsona wnioski normatywne nie wynikają też z moich eksplikacji pojęciowych, które są pod wskazanym względem całkowicie neutralne.

5.3. Trzeci zarzut

Na gruncie rekonstrukcji kolejnego zarzutu Strawsona kierowanego przeciwko tradycyjnemu kompatybilistycie, krytyka i uzasadnienia sądów o odpowiedzialności nie mogą mieć charakteru archimedesowego: zarówno uzasadnienie, jak i krytyka sądów o odpowiedzialności podmiotów ma i może mieć charakter jedynie wewnętrzny. Każdy optymista, który chciałby np. twierdzić, że przypisywanie odpowiedzialności jest uzasadnione jedynie w tych przypadkach, w których może przyczynić się do podtrzymania porządku normatywnego, zajmuje stanowisko w ramach tradycyjnej instytucji przypisywania odpowiedzialności – nie zaś spoza niej¹⁷.

¹⁵ Strawson 1993, s. 62, 63–64.

¹⁶ Strawson 1993, s. 62.

¹⁷ Strawson 1993, s. 64, 66.

Przedstawione analizy pojęciowe dobrze oddają fakt wskazywany przez Strawsona. Ktoś, kto uważa, że odpowiedzialność nie powinna być przypisywana tam, gdzie jest to nieużyteczne lub nieskuteczne w podtrzymywaniu pożądanego porządku społecznego, dokonuje asercji sądu na temat okoliczności wyłączających odpowiedzialność. Osoba taka planuje, by nie odczuwać moralnego oburzenia wobec tych podmiotów, których motyw nie są wrażliwe na manifestacje moralnego oburzenia otoczenia, oraz wobec tych, których zachowanie może być efektywniej kontrolowane w inny sposób. Stanowisko takie jest wyrażalne w aparacie pojęciowym właściwym dla praktyki przypisywania odpowiedzialności przy założeniu trafności mojej analizy tego aparatu, choć – oczywiście – z samej zaprezentowanej analizy pojęciowej nie wynika.

6. Janusowe oblicze sporu o kompatybilizm

W świetle przedstawionych analiz pojęciowych spór o warunki odpowiedzialności moralnej jest sporem należącym wyłącznie do etyki normatywnej. Nie jest zatem prawdą, że jakkolwiek okoliczność *O* jest taka, że stwierdzenie, że sprawca jest odpowiedzialny moralnie w *O*, jest wykroczeniem przeciwko pojęciu odpowiedzialności moralnej. *A fortiori* dotyczy to okoliczności, w których działania podmiotu są zdeterminowane. Jeśli zatem spór o kompatybilizm interpretować jako spór o **pojęciową możliwość** odpowiedzialności w deterministycznym świecie, to w świetle powyższych analiz i Strawsonowskich argumentów rację należy przyznać kompatybilistom.

Wzmiankowany spór może być jednak interpretowany inaczej, tzn. jako spór o to, czy faktycznie uznawane przez nas okoliczności znoszące odpowiedzialność moralną są takie, że jeśli nasz świat jest deterministyczny, to jest prawdą, że wszyscy działamy w tego rodzaju okolicznościach. Kłopot w tym, że taka konceptualizacja sporu o kompatybilizm czyni pogląd przeciwny jawnie absurdalnym. Wyraźnie wynika to z rozważań Strawsona nad obecnymi w naszej praktyce warunkami odpowiedzialności. Nawet jeśli nasz świat jest deterministyczny, to przecież nie jest prawdą, że wszyscy jesteśmy podmiotami działającymi, których pogląd na świat jest urojeniem szaleńca lub których czyny nie mogą być zrozumiane w kategoriach świadomych zamiarów, że wszyscy jesteśmy neurotykami lub dziećmi, że wszyscy nie wiemy, co czynimy, lub nie chcemy tego, co wywołujemy w świecie naszymi działaniami itd.

Przedstawione powyżej dwie możliwe interpretacje sporu o kompatybilizm nie wyczerpują zbioru wiarygodnych ujęć. Nie stoją one w sprzeczności ze stanowiskiem, zgodnie z którym wprawdzie odpowiedzialność w świecie deterministycznym jest pojęciowo możliwa, a faktycznie stosowane przez nas warunki odpowiedzialności mają naturę kompatybilistyczną, to jednak odpo-

wiedzialność moralna w deterministycznym świecie nie jest możliwa metafizycznie. Jeśli spór o kompatybilizm interpretować jako spór o **metafizyczną możliwość**, to cała dyskusja uzyskuje charakter substancjalny i jako taka nie podlega rozstrzygnięciu na drodze czysto pojęciowej analizy; nie jest też już to spór o to, jakie są praktycznie stosowane przez nas warunki odpowiedzialności, lecz o to, jakie warunki te być powinny. Jest wielką zaletą ekspresywistycznej wykładni pojęć moralnych, w tym pojęcia odpowiedzialności, że tworzy pojęciową przestrzeń dla takiej interpretacji sporu o kompatybilizm.

W ramach tej interpretacji antykompatybilista zajmuje stanowisko z dziedziny etyki normatywnej. Zgodnie z nim, zbiór okoliczności znoszących odpowiedzialność zawiera okoliczność bycia zdeterminowanym lub zawiera takie okoliczności, że jeśli świat jest deterministyczny, to w świecie tym nie ma podmiotów, które czyniąc coś złego, działałyby w innych warunkach niż te należące do wzmiankowanego zbioru. W konsekwencji pesymista, inaczej niż optymista, będzie utrzymywał, że wszystkie sądy przypisujące odpowiedzialność podmiotom działającym w świecie deterministycznym są pozbawione uzasadnienia. Niemniej jednak pesymista nie będzie mógł poprzeć swojego poglądu w tej sprawie jakimkolwiek poprawnym rozumowaniem wychodzącym z przesłanek ograniczających się do eksplikacji pojęcia odpowiedzialności oraz treści teorii determinizmu. Pesymista i optymista mogą zgadzać się ze sobą co do analizy pojęcia odpowiedzialności i treści doktryny determinizmu, lecz nadal różnić się w kwestii odpowiedzialności podmiotów zdeterminowanych. Aby sygnalizowane rozumowanie pesymisty było poprawne, konieczne będą przesłanki dodatkowe, których cechą istotną będzie charakter normatywny – będą to substancjalne sądy moralne. Argumenty krytyczne optymisty będą z kolei zmierzać do podważenia tych dodatkowych przesłanek.

Jeśli spór kompatybilistów z ich przeciwnikami interpretować jako spór normatywny o metafizyczną możliwość, to same te stanowiska stają się stanowiskami normatywnymi. Oczywiście twierdzenie, że odpowiedzialność w świecie deterministycznym jest lub nie jest metafizycznie możliwa, jest bardzo ogólne. Dlatego konkretne teorie kompatybilistyczne, podobnie jak antykompatybilistyczne, powinny zawierać więcej treści. W tym kontekście możemy pomyśleć o możliwej radykalnej wersji normatywnej koncepcji kompatybilistycznej. Mogłaby ona stanowić proste przekształcenie hipotezy funkcjonalistycznej, wyjaśniającej praktykę przypisywania odpowiedzialności moralnej, w teorię normatywną dotyczącą warunków odpowiedzialności. Głosiłaby ona, że odpowiedzialność moralną ponoszą tylko takie jednostki, których ukaranie lub uczynienie przedmiotem uczuć reaktywnych jest skutecznym instrumentem utrzymywania porządku normatywnego. Niezależnie od tego, co powinniśmy myśleć o trafności takiej propozycji, podkreślmy, że ta wersja kompatybilizmu nie wynika ani z analizy pojęcia odpowiedzialności w koniunkcji z hipotezą

funkcjonalistyczną (8), ani z ogólnego twierdzenia kompatybilizmu w ujęciu normatywnym, zgodnie z którym istnieją metafizycznie możliwe światy deterministyczne, w których podmioty działające są odpowiedzialne za swoje czyny.

7. Podsumowanie

Peter Strawson ma rację, gdy krytykuje to, co uważa za tradycyjną wersję stanowiska kompatybilistycznego, i nie ma słuszności, gdy próbuje sprowadzić pogląd antykompatybilistyczny do niedorzeczności. O tym ostatnim przekonuje argument odwołujący się do świata mechanicznych pomarańczy. To wszystko nie oznacza jednak, że powinniśmy ogłosić triumf pesymisty – wręcz przeciwnie. Połączenie ekspresywistycznych analiz pojęciowych z kompatybilistyczną hipotezą funkcjonalistyczną pozwala na uwolnienie kompatybilizmu od zarzutów Strawsona. Co więcej, droga, która do tego uwolnienia prowadzi, dobrze wpasowuje się w rozważania Strawsona nad istotnością moralnych uczuć i postaw reaktywnych dla praktyki przypisywania odpowiedzialności moralnej; wpisuje się także w zasadniczy kierunek argumentacji autora *Freedom and Resentment* – argumentacji wspierającej optymiste. Spór o pojęciową możliwość odpowiedzialności w świecie deterministycznym wygrywa kompatybilista. Z tego faktu nie wynika jednak, że wygrywa on także dyskusję nad możliwością metafizyczną. Nawet jeśli i tutaj racja jest po stronie kompatybilisty, to wykazanie tego wymaga zaangażowania czegoś więcej niż analiz pojęciowych – wymaga zajęcia stanowiska w dziedzinie etyki normatywnej.

Bibliografia

- Burgess A. (2017), *Mechaniczna pomarańcza*, przeł. R. Stiller, Vis-à-vis Etiuda, Kraków.
- Gibbard A. (1990), *Wise Choices, Apt Feelings. A Theory of Normative Judgment*, Harvard University Press, Cambridge, Mass.
- Nadelhoffer T., Gromet D., Goodwin G., Nahmias E., Sripada C., Sinnott-Armstrong W. (2013), *The Mind, the Brain, and the Law*, w: T.A. Nadelhoffer (ed.), *The Future of Punishment*, Oxford University Press, Oxford – New York, s. 193–211.
- Searle J.R. (1985), *Intentionality. An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge.
- Stiller R. (2017), *Kilka sprężyn z nakręcanej pomarańczy*, w: A. Burgess, *Mechaniczna pomarańcza*, przeł. R. Stiller, Vis-à-vis Etiuda, Kraków, s. 197–226.

- Strawson P. (1993), *Freedom and Resentment*, w: J.M. Fischer, M. Ravizza (eds.), *Perspectives on Moral Responsibility*, Cornell University Press, Ithaca – London, s. 45–66.
- Williams B. (1999), *Traf moralny*, w: tenże, „*Ile wolności powinna mieć wola?*” i inne eseje filozoficzne, Fundacja Aletheia, Warszawa, s. 215–238.

A d r i a n K u ż n i a r

Compatibilism and the objective attitude in the world of clockwork oranges

Keywords: *compatibilism, objective attitude, reactive moral sentiments, moral responsibility, expressivism*

The paper focuses on two research objectives. First, it aims to critically examine a *reductio ad absurdum* argument against incompatibilism whose main themes can be found in Peter F. Strawson's *Freedom and Resentment*. The doubts raised about the argument are inspired by a thought experiment based on fictitious Ludovico's technique described in Anthony Burgess's novel *A Clockwork Orange*. The second objective consists in outlining a version of the compatibilist stance – the version which is immune to Strawson's objections against the traditional rendering of compatibilism and enables deeper understanding of various possible interpretations of the controversy between compatibilists and their opponents. The proposed position includes a hypothesis on the function of the attitude of participation and the expressivist explications of the concepts crucial for the practice of ascribing moral responsibility. The important feature of the analyses in question is the central role of the states of mind whose content are plans for reactive moral sentiments.