

K a r o l M i l c z a r e k

Alfred Mele – dwie ekwiwokacje w dyskusji o słabości woli

Słowa kluczowe: *akrazja, A. Mele, pragnienie, racjonalność, słaba wola*

Problem słabości woli (akrazji¹) współcześnie przybrał formę pytania o specyficznego rodzaju działania. Są to przejawy słabości woli, określane także jako działania niepowściągliwe czy też akratyczne, zwykle definiowane jako wolne, intencjonalne działania wbrew najlepszemu sądowi. Jedną z aktualnie najbardziej oddziałujących koncepcji mających wyjaśnić naturę tych działań proponował amerykański filozof Alfred Mele, który w książkach *Irrationality* (Mele 1987), *Autonomous Agents* (Mele 1995) oraz *Backsliding* (Mele 2012), a także licznych artykułach poświęconych miejscu słabości woli w teorii działania intencjonalnego przedstawił oryginalną teorię działania akratycznego.

Na wstępie warto odpowiedzieć na pytanie, dlaczego słabość woli miałaby stanowić wyzwanie dla filozoficznej teorii działania. Jest tak, ponieważ – jak wskazuje na poziomie metateoretycznym sam Mele – ludzkie działanie można analizować z dwóch perspektyw: motywacyjnej oraz intelektualnej. Pierwszą daje się zastosować do wszystkich organizmów żywych i jest stosowana w badaniach empirycznych, gdyż pozwala na umieszczenie działań w porządku naturalnych przyczyn. Jeśli byłby to jedyny sposób opisu, wyjaśnienie działań

Karol Milczarek, Uniwersytet Warszawski, Wydział Filozofii, ul. Krakowskie Przedmieście 3, 00-047 Warszawa; e-mail: kj.milczarek3@uw.edu.pl, ORCID: 0000-0002-0497-7252.

Dziękuję Michałowi Barczowi i Marcie Zarębie za wszystkie komentarze i uwagi do wcześniejszych wersji tego tekstu.

¹ Pomijam tu skądinąd interesującą dyskusję na temat zasadności utożsamienia klasycznego pojęcia akrazji z potocznym określeniem słabości woli – a właściwie z angielskim *weakness of will* / *weak will* (zob. Mele 2010; May, Holton 2012).

akratycznych nie stanowiłoby problemu, ponieważ wystarczyłoby uznać, że sprawca był po prostu bardziej zmotywowany do podjęcia działania, którego nie uważa za najlepsze. Między pewnym działaniem a oceną, że najlepiej będzie się od niego powstrzymać, nie ma ani logicznej, ani nomologicznej sprzeczności. Jednak druga perspektywa, wywodząca się z potocznej praktyki wyjaśniania działań poprzez przypisywanie sprawcom stanów umysłowych, a w zachodniej filozofii obecna w zasadzie od momentu jej narodzin, bo od czasów rozpoczęcia myślenia o człowieku jako o zwierzęciu racjonalnym, włącza w dyskusję pytanie o racje. Stosuje się ona do podmiotów racjonalnych, których działania wyjaśniane są przez żywione nastawienia psychologiczne o treści propozycyjalnej (tzw. nastawienia sądeniowe). Z tej perspektywy działania niepowściągliwe są *prima facie* niemożliwe, bowiem każdemu (a przynajmniej każdemu wolnemu) działaniu powinien odpowiadać kierujący działaniem osąd praktyczny, głoszący że to właśnie działanie jest najlepsze do podjęcia. Dzięki przypisaniu takiego osądu, działanie sprawcy staje się zrozumiałe – umieszczone w porządku racji. Na przecięciu obu tych perspektyw rodzi się problem słabości woli (Mele 2012, s. 60–64).

Wyjaśnienie słabości woli stanowi szczególne wyzwanie dla zwolenników przyczynowej teorii działania intencjonalnego, ponieważ – zgodnie z klasycznym Davidsonowskim ujęciem (Davidson 1963/1980) – na jej gruncie racje (pragnienia i przekonania) służące do wyjaśniania działań są jednocześnie ich przyczynami². W wypadku teorii postulujących większą różnorodność zjawisk mentalnych niezbędnych do sformułowania adekwatnego przyczynowego wyjaśnienia działania intencjonalnego (np. uwzględniających najlepsze sądy oraz intencje, a także zdolność samokontroli, jak głosi m.in. autorska koncepcja Melego) obecny jest ten sam problem, gdyż wszystkie ogniwa pośredniczące między racjami sprawcy a jego zachowaniem także pozostają w dwoistym – przyczynowym i logicznym – związku z działaniem (Mele 1992).

Tak postawione wyzwanie stanowi tło dla zasadniczego celu artykułu, jakim jest omówienie dwóch ekwiwokacji obecnych w dyskusji nad słabością woli, a szczególnie istotnych w kontekście krytycznej analizy stanowiska Melego, zrekonstruowanego w § 1. Pierwsza, dotycząca pojęcia pragnienia (używanego albo w sensie ewaluacyjnym, albo motywacyjnym), pozwala mu wykazać niepoprawność zarzutów sceptycznych – negujących możliwość działania ze słabej woli (§ 2). Druga, odnosząca się do pojęcia racjonalności (rozumianej albo jako opisywalność w kategoriach intencjonalnych, albo jako wewnętrzna spójność sprawcy), daje się znaleźć w samej argumentacji Me-

² Donald Davidson przedstawił szkicową propozycję rozwiązania problemu słabości woli w artykule *How Is Weakness of the Will Possible? (Jak jest możliwa słabość woli?)* (Davidson 1970/1997), której omówienie można znaleźć m.in. w: Lazar 1996; Machciewicz-Grad 2020. Także koncepcję Melego można potraktować jako głos w dyskusji ze stanowiskiem Davidsona.

lego i świadczy o nieadekwatności jego koncepcji – jest tak względu na fakt, że nie dostarcza ona wystarczających podstaw dla uznania działań akratycznych za nieracjonalne (§ 3).

1. Stanowisko Alfreda Melego: rekonstrukcja

Podstawową tezę wysuniętą przez Melego jest to, że słabość woli faktycznie istnieje, więc *a fortiori*, że działania akratyczne są psychologicznie możliwe. Sformułowanie adekwatnej koncepcji słabej woli ma służyć między innymi jako argument przemawiający za tą tezą (Mele 2012, s. 116–121). Wbrew koncepcjom rewizyjnym, postulującym alternatywne eksplikacje słabości woli³, Mele jedynie uściśla klasyczne pojęcie akrazji. Działanie ściśle niepowściągliwe definiuje następująco:

Działanie *A* jest działaniem ściśle niepowściągliwym wtedy i tylko wtedy, gdy jest wykonywane intencjonalnie i w sposób wolny oraz, w czasie, gdy jest wykonywane, jego sprawca świadomie utrzymuje sąd, z którego wynika, że ma on dobry i wystarczający powód dla niewykonywania *A* w tym czasie (Mele 1987, s. 7).

W późniejszej pracy dodaje, że sprawca powinien być w dobrej kondycji psychicznej (zdrowy umysłowo, bez depresji), by odróżnić akrazję od innych źródeł irracjonalności (Mele 2012, s. 8–10), a także podkreśla, że sąd, wbrew któremu on działa, ma być sądem praktycznym – sformułowanym w trakcie deliberacji praktycznej (Mele 2012, s. 14).

Ów sąd, określany często mianem najlepszego bądź lepszego sądu (*best/better judgement*), to dla Melego specyficzny rodzaj przekonania⁴. Uwzględnia ono wszystkie dostępne sprawcy racje, więc także racje moralne, ale nie tylko je – w pewnych sytuacjach ktoś może bowiem uznać, że najlepiej będzie przedłożyć osobistą korzyść nad swoje powinności. Funkcjonalną rolę najlepszego sądu – ze względu na jego źródłowy związek z deliberacją praktyczną –

³ Do koncepcji rewizyjnych należą – by podać kilka przykładów – koncepcje eksplikujące słabość woli jako: (a) działanie niezgodne z pragnieniami drugiego (lub wyższego) rzędu (Jeffrey 1974; Bigelow i in. 1990; Janikowski 2016), (b) działanie wywołane pragnieniem niepodatnym lub nie dość podatnym na informacje (Jackson 1984; Michael 2019), (c) działanie zbliżone do działania rutynowego – umiejętnego, ale niepoprzedzonego deliberacją (Burch 2018), (d) działanie wbrew uświadomionej racji (Johnson 2021), (e) nadmierną skłonność do rewizji powziętych rezolucji (Holton 1999, 2003).

⁴ Dla podkreślenia praktycznego charakteru najlepszych sądów niekiedy określa on je jako przekonania typu B, pod pewnymi względami różne od zwyczajnych przekonania typu A (Mele 1991, s. 43). Uzasadnione jest w tym kontekście pytanie, czy różnice między tymi dwoma typami przekonania nie są zbyt duże, by wciąż traktować je jako przynależące do jednej kategorii stanów mentalnych, ale pomijam to w niniejszym tekście.

jest kierowanie działaniem. Jednocześnie stanowi on (subiektywne) rozstrzygnięcie pytania, co najlepiej jest zrobić, i z tego powodu zobowiązuje on sprawcę do podjęcia określonego działania. W tym sensie najlepsze sądy są źródłem zobowiązań ewaluacyjnych (Mele 1995, s. 15–16, 69–76).

Sformułowanie adekwatnej koncepcji słabej woli wymaga jeszcze wyjaśnienia, jak to możliwe, że sprawcom zdarza się działać wbrew tak scharakteryzowanym najlepszym sądom, a także dlaczego takie działanie może zostać opisane jako intencjonalne oraz wolne. Według Melego do takiego wniosku prowadzą następujące przesłanki:

- (1) Nasze najlepsze sądy normalnie są przynajmniej częściowo oparte na ewaluacji przedmiotów naszych pragnień (tj. pożądanых rzeczy).
- (2) Siła motywacyjna naszych pragnień nie zawsze odpowiada ewaluacji przedmiotów naszych pragnień (Mele 2012, s. 119).

Wraz z założeniem, że sprawca wykonuje to działanie, do którego jest najbardziej zmotywowany, prowadzą one do wniosku, że niekiedy zdarza nam się działać wbrew naszym najlepszym sądom. Dalej, aby rzucić nieco więcej światła na przesłankę (1), należy przejść do zagadnienia intencjonalności działań akratycznych, a na przesłankę (2) – do zagadnienia ich wolności.

Rozważając typowe przykłady słabości woli – picie alkoholu, objadanie się, palenie, wdawanie się w romanse czy zaniedbywanie obowiązków na rzecz wypoczynku – łatwo zaobserwować, że sprawcy działają akratycznie bynajmniej nie dlatego, że uważają to działanie za mniej wartościowe (co skądinąd oczywiście jest prawdą), ale że – pod pewnym względem – jest ono dla nich atrakcyjne. Zdaniem Melego oznacza to, że działanie akratyczne może być opisane jako intencjonalne, gdyż – tak jak w przypadku wszystkich pozostałych działań – za przejawami słabości woli stoją pewne racje, zwykle dotyczące krótkotrwałej czy osiągalnej w najbliższej przyszłości przyjemności. Działanie ze słabej woli stanowi więc satysfakcję pragnienia sprawcy, które (wraz z odpowiednim przekonaniem) stanowi rację za podjęciem działania, jednocześnie uzasadniającą, jak i powodującą (za pośrednictwem intencji) działanie⁵. Racja ta – w odróżnieniu od racji, w oparciu o które sprawca sformułował najlepszy sąd – staje się racją efektywną, czyli przyczynowo odpowiedzialną za spowodowanie działania (por. Mele 1992, rozdz. 6–7).

Zatem przy słabości woli możliwe jest udzielenie odpowiedzi na pytanie: „Dlaczego (sprawca postąpił w dany sposób)?”, czyli wskazanie powodu (ra-

⁵ Dla zachowania ścisłości należy dodać, że w innym kontekście Mele deklaruje neutralność co do ontologii racji, przy czym twierdzi, że jeśli racje są przedmiotami abstrakcyjnymi (np. faktami), to do wyjaśnienia działań istotne są psychologiczne nastawienia wobec tychże abstraktów, co w konsekwencji sprowadza wyjaśnienie z powrotem w dziedzinę stanów mentalnych (Mele 1992, s. 104–105).

cji) działania, którego obecność klasycznie uznaje się za wyznaczającą linię demarkacyjną między działaniami a niezamierzonymi zachowaniami (Anscombe 1957; Davidson 1963/1980). Z tego powodu, jak twierdzi Mele, działanie akratyczne jest minimalnie racjonalne (Mele 1987, s. 46–48, 96–101), tak jak minimalnie racjonalne jest każde działanie, aby w ogóle mogło zostać opisane jako działanie. Jednocześnie przedmiotami pragnień włączonych w te racje są te stany rzeczy, które według oceny samego sprawcy (wyrażonej w najlepszym sądzie) nie są najbardziej warte realizacji, co odróżnia działania akratyczne od pozostałych działań.

Poza intencjonalnością, bardziej szczegółowego wyjaśnienia domaga się także odróżnienie słabości woli od przymusu, a więc wskazanie, dlaczego działanie akratyczne jest wolne. Zdaniem Melego kluczem do wyjaśnienia możliwości działania akratycznego jest wykorzystanie pojęć siły motywacyjnej i samokontroli w opisie etiologii działania intencjonalnego. Siła motywacyjna, rozumiana w przybliżeniu jako „moc wywołująca działania” (Mele 1992, s. 51), odnosi się do konatywnego wymiaru pragnienia (lub innego motywacyjnego efektywnego nastawienia sądzeniowego). Siła danego pragnienia zależy od jego pozytywnej i negatywnej bazy motywacyjnej, łącznie konstytuujących całościową bazę motywacyjną⁶, dodatkowo zmodyfikowanej przez tzw. czynniki atrakcyjności (Sripada 2014, s. 8) o charakterze kognitywnym: m.in. sposób reprezentacji przedmiotu pragnienia, nakierowanie uwagi sprawcy czy przewidywany do uzyskania gratyfikacji czas (Mele 1987, s. 67–69).

Samokontrola to natomiast zdolność (lub – gdy mowa o aktach podjęcia samokontroli – manifestacja tej zdolności) do wywierania wpływu na swoją kondycję motywacyjną, paradygmatycznie: tak aby pozostawała w zgodzie z własną perspektywą ewaluacyjną (wyrażaną w najlepszych sądach) lub egzekucyjną (wyrażaną w intencjach). Pozwala ona więc sprawcy przeciwstawić się pokusie działania akratycznego (Mele 2003, s. 3). Ujawnia się, gdy domyślna ścieżka formułowania intencji na bazie praktycznego sądu zawodzi. Mele rozważa trzy warianty: (a) normalny, gdy wszystkie stany mentalne

⁶ Pozytywna baza motywacyjna jest tu rozumiana jako zbiór wszystkich komponentów motywacyjnych zwiększających siłę motywacyjną danego pragnienia, natomiast negatywna baza motywacyjna to, analogicznie, zbiór wszystkich komponentów zmniejszających tę siłę (Mele 2003, s. 19–20). Rozróżnienie czynników „popychających” do i „odpychających” od podjęcia działania pozwala na udzielenie odpowiedzi na dwa pytania, wyrażające coś, co autor określa jako paradoksy samokontroli: (a) Dlaczego sprawca, który w chwili t_1 jest bardziej zmotywowany do podjęcia działania α w późniejszej chwili t_2 , również w t_1 może być bardziej zmotywowany do podjęcia samokontroli, by nie wykonać α w t_2 ? (b) Dlaczego sprawca nie jest po prostu bardziej zmotywowany do wykonania działania α w chwili t_2 , jeśli jest najbardziej zmotywowany do podjęcia samokontroli we wcześniejszym t_1 , by podjąć pewne działanie α w t_2 ? (zob. Mele 1987, s. 62–74).

spełniają swoje założone funkcje, a w konsekwencji wydanie sądu praktycznego bezproblemowo (bez udziału samokontroli) powoduje sformułowanie odpowiedniej intencji, (b) pośredni, w którym pojawienie się innego pragnienia zakłóca domyślny związek, ale sprawca i tak formułuje odpowiednią intencję, (c) akratyczny, gdzie sprawca ulega pragnieniu (mimo że miał możliwość oparcia mu się). Istnienie sytuacji (b), a także odróżnienie (c) od przymusu wewnętrznego, kiedy to sprawca działa ze względu na zniewalające pragnienie, gwarantuje właśnie samokontrola (s. 27–28)⁷.

Struktura działania akratycznego przedstawia się zatem następująco: gdy najlepszy sąd nie spełnia przypisanej mu roli funkcjonalnej, a więc nie inicjuje i nie kieruje postępowaniem sprawcy, wówczas następuje przerwanie domyślnej ścieżki generowania działania intencjonalnego. Zdaniem Melego, o ile psychologiczna konstrukcja człowieka sprawia, że zwykle najlepszy sąd, by wykonać pewne działanie α , jest przyczynowo oraz treściowo powiązany z intencją do α -owania, ta natomiast (jako intencja proksymalna) jest bezpośrednią przyczyną α -owania (Mele 1992, s. 228–241; Mele 2003, s. 25–30), to możliwa jest sytuacja, gdy inne pragnienie – oceniane jako mniej godne realizacji, ale silniejsze motywacyjnie – prowadzi do sformułowania innej intencji proksymalnej, by nie α -ować. Najlepszy sąd sprawcy może wówczas pozostać niezmieniony, gdyż żadne istotne fakty na temat wartości uznawanych przez sprawcę nie uległy zmianie (Mele 2012, s. 84). Jeśli wówczas sprawca posiadał zdolność podjęcia samokontroli, lecz nie podjął próby jej wykorzystania – ze względu na fakt, że: (a) wybrał zaniechanie podjęcia samokontroli (co określić można jako pobłażanie sobie), (b) nie uwzględnił w deliberacji konieczności podjęcia samokontroli lub (c) niewłaściwie ocenił wysiłek i rodzaj samokontroli konieczny do oparcia się pokusie (Mele 1987, s. 27–28) – podejmuje on działanie akratyczne.

2. Pierwsza ekwiwokacja: pojęcie pragnienia

Aby pokazać, w jaki sposób dochodzi do działania akratycznego, Mele opisuje kondycję motywacyjną sprawcy jako do pewnego stopnia niezależną od wyniku jego deliberacji praktycznej. Związek najlepszych sądów z działaniem jest bowiem systematyczny, ale nie bezwyjątkowy. Zdaniem filozofów sceptycznych wobec możliwości działania akratycznego (zob. np. Watson 1977; Pugmire 1982; Tenenbaum 1999)⁸ – w uproszczonej wersji ich argumentacji –

⁷ Dyskusję na temat pewnych spornych punktów w charakterystyce samokontroli prowadzi z autorem Chandra Sripada (Sripada 2014; por. odpowiedź w: Mele 2014).

⁸ W dalszej części artykułu omówione są poszczególne argumenty tych autorów.

z przyjęcia takiego założenia wynikałoby paradoksalne twierdzenie, że sprawca może dobrowolnie uczynić coś, nawet jeśli nie chciałby tego robić.

Mele rozróżnia jednak dwa sensy, w jakich można mówić o pragnieniach (czy też chceniach) podjęcia danego działania: motywacyjny oraz ewaluacyjny. Chcieć czegoś, czyli mieć jakieś pragnienie, w swoim bazowym sensie oznacza posiadanie motywacji do wykonania pewnego działania (Mele 1992, s. 47). Jednak stwierdzenie, że ktoś chce wykonać działanie α bardziej niż β , nie zawsze dotyczy kondycji motywacyjnej sprawcy. Może bowiem odnosić się także do tego, które działanie sprawca ocenia jako lepsze, a więc jaką relatywną wartość przypisuje przedmiotom swoich pragnień, co stanowi użycie pojęcia pragnienia w sensie ewaluacyjnym. Pojęcie to jest więc dwuznaczne (Mele 1987, s. 37).

Tylko w pierwszym z wyróżnionych znaczeń pragnienie jest ściśle powiązane z motywacją – trywialną prawdą jest stwierdzenie o związku pragnień w sensie motywacyjnym (pragnień_m) z motywacją (M)⁹:

(P_mM) Jeśli sprawca S najbardziej pragnie_m wykonać działanie α , to jest najbardziej zmotywowany do α -owania (por. Mele 1987, s. 13–14).

Czym innym jest jednak analogiczna teza dla pragnień w sensie ewaluacyjnym (pragnień_e):

(P_eM) Jeśli sprawca S najbardziej pragnie_e wykonać działanie α , to jest najbardziej zmotywowany do α -owania.

Jako że za wyraz najsilniejszego pragnienia w sensie ewaluacyjnym można uznać najlepszy sąd (NS), z dwóch powyższych wyłącznie teza (P_eM) jest równoważna z tezą:

(NSM) Jeśli sprawca S żywi najlepszy sąd, by α -ować, to S jest najbardziej zmotywowany do α -owania.

Rozróżnienie motywacji i ewaluacji wprowadza konceptualną przestrzeń dla odrzucenia sceptycznych argumentów bazujących jedynie na (P_mM), a dla których poprawności w istocie wymagana jest prawdziwość (NSM)¹⁰. Widać to dobrze na przykładzie przywołanego wyżej twierdzenia, że sprawca α -uje,

⁹ Dla zachowania prostoty, w poniższych tezach pomijam indeksy czasowe.

¹⁰ Przy czym należy mieć na uwadze, że w tym momencie argumentacyjnej dyalektyki nie zostało jeszcze powiedziane nic na rzecz fałszywości (NSM). Dodatkowo warto zauważyć, że fałszywość (NSM) jest praktycznie tożsama z prawdziwością przesłanki (2) argumentu przywołanego w § 1, gdyż, przy założeniach: (a) synonimiczności najlepszego sądu z ewaluacją przedmiotów pragnień, (b) uznania, że źródłem motywacji do działania są pragnienia (por. Mele 1989), są to tezy logicznie sprzeczne – zatem każdy argument przeciwko (NSM) będzie stanowił rację na rzecz (2) i odwrotnie.

ale (najbardziej) chciałby powstrzymać się od α -owania. O ile w mowie potocznej mogą zdarzyć się konteksty, w których nie brzmi ono dewiacyjnie, w istocie głosi ono, że w tym samym czasie sprawca żywi dwa różne „najsilniejsze” pragnienia. Eksplikacja takich wyrażen z użyciem rozróżnienia tez ($P_m M$) i ($P_e M$) pozwala na uniknięcie tej konsekwencji – zależnie od kontekstu zdanie to może wyrażać sąd, że sprawca jest najbardziej zmotywowany do α -owania, ale nie ocenia α -owania jako najlepszego działania, albo że według sprawcy najlepiej byłoby α -ować, ale nie jest on wystarczająco zmotywowany, by faktycznie α -ować.

Ekwiwokacja ta daje się wychwycić w trzech argumentach sceptycznych wobec tezy o istnieniu słabej woli, które – za Arthurem Walkerem – można scharakteryzować w następujący sposób:

- (I) Argument ze zniewolenia (*Enslavement Argument*): żadne działanie *prima facie* akratyczne nie jest działaniem wolnym. Zatem – jeśli definicyjnie słabość woli zakłada wolność (rozumianą zwykle jako zdolność sprawcy do wykonania innego działania niż to, które faktycznie wykonał) – w rzeczywistości nie istnieją żadne działania naprawdę akratyczne (Walker 1989, s. 654–658).
- (II) Argument z braku intencjonalności (*No Intentional Action Argument*): żadne działanie *prima facie* akratyczne nie jest intencjonalne, ponieważ: (a) sprawca nie ma powodu do działania niepowściągliwego, (b) działanie akratyczne nie egzemplifikuje odpowiedniej inferencji praktycznej, (c) intencjonalne α -owanie pociąga ocenianie α jako najlepszego z możliwych działań, jak głosi argument „ze związku wstecz” (*Backward Connection Argument*). Zatem – jeśli definicyjnie słabość woli zakłada intencjonalność – nie istnieją żadne działania naprawdę akratyczne (Walker 1989, s. 658–666).
- (III) Argument „ze związku wprzód” (*Forward Connection Argument*): ewaluacja działania jest ściśle związana z motywacją do działania. Zatem – jeśli definicyjnie słabość woli zakłada niezgodność najlepszego sądu z działaniem, a dla tej konieczna jest rozbieżność ewaluacji z motywacją – nie istnieją żadne działania naprawdę akratyczne (Walker 1989, s. 666–669).

Pierwszy z tych argumentów przyjmuje różne formy. David Pugmire pyta, co więcej sprawca mógłby uczynić poza deliberacją i sformułowaniem najlepszego sądu (Pears, Pugmire 1982, s. 189). Gary Watson wskazuje, że odwołanie się do samokontroli jest wystarczające dla uznania działania akratycznego jako wolnego, gdyż adekwatna koncepcja słabości woli powinna także dostarczyć wyjaśnienia, dlaczego sprawca nie podjął samokontroli, jeśli był do tego zdolny – a jeśli nie zrobił tego ze względu na fakt, że wybrał niepodjęcie samokontroli lub że źle ocenił wysiłek potrzebny do oparcia się pokusie (zdaniem Watsona te opcje nie pozwalają na wyjaśnienie działania jako akratycz-

nego), musi to oznaczać, że nie mógł się oprzeć pragnieniu, a więc że nie był zdolny do podjęcia wystarczającej samokontroli (Watson 1977, s. 333–338). Sergio Tenenbaum podkreśla natomiast, że aby uzasadnić wolność działań akratycznych, nie można odwołać się do wolnego wyboru, by nie oprzeć się danej pokusie, ponieważ oznaczałoby to albo, że sprawca w istocie oceniał działanie akratyczne jako bardziej wartościowe, albo że sam ten wybór był akratyczny, co stanowi jedynie przesunięcie problemu na niższy poziom – wolności akratycznego wyboru (Tenenbaum 1999).

Mele wykorzystuje rozróżnienie dwóch płaszczyzn – ewaluacyjnej i motywacyjnej – w odpowiedzi na każdy z tych zarzutów. W odniesieniu do pytania Pugmire’a wskazuje, że istnieją niedeliberowane metody samokontroli – wpływające na kondycję motywacyjną sprawcy, ale niewpływające w żaden sposób na ewaluację działania (Mele 1987, s. 51–55). Ponadto neguje założenie Watsona, że pobłażanie sobie lub niewystarczający wysiłek wyklucza akrazję, i dodaje, że są też takie przypadki, w których sprawca w ogóle nie wydaje sądu co do tego, jaki stopień czy metoda samokontroli byłyby w danej sytuacji adekwatne (Mele 1986, s. 677; Mele 1987, s. 26). W dwóch ostatnich wypadkach za niewystarczającą motywację sprawcy do podjęcia działania zgodnego z najlepszym sądem nie odpowiada żadna zmiana w ewaluacji działania, lecz „błędy techniczne” – nietrafny sąd na temat samokontroli lub brak takiego sądu. Natomiast Tenenbaumowi (w kontekście scenariusza pobłażania sobie) odpowiada, że nie jest prawdą, iż akratyczny wybór niepodjęcia samokontroli należałoby znów wyjaśnić przez odwołanie do akratycznego wyboru podjęcia akratycznego wyboru (lub innego niepowściągliwego działania). Według Melego w przypadkach pobłażania sobie coś może odpowiadać za akratyczny wybór sprawcy, a jednocześnie samo w sobie nie być mentalnym działaniem niepowściągliwym (Mele 2012, s. 42–43). Ową przyczyną, a właściwie różnymi przyczynami, są bowiem nieewaluacyjne czynniki wpływające na bazę motywacyjną danego działania, w szczególności czas przewidywany do gratyfikacji czy sposób nakierowania uwagi sprawcy (Mele 1987, s. 93–95).

O ile człony (IIa) i (IIb) argumentu z braku intencjonalności nie wymagają rozróżnienia dwóch sensów pragnienia – zgodnie z koncepcją Melego działanie akratyczne jest bowiem działaniem ze względu na racje, więc istnieje tu jakaś racja do działania oraz sprawca postępuje zgodnie (a nie jedynie w przypadkowej zbieżności) z nią – to rozróżnienie takie zostaje wykorzystane do odparcia twierdzenia, że jeśli *S* intencjonalnie α -uje, to *S* żywi najlepszy sąd, by α -ować, na którym oparty jest argument (IIc).

David F. Pears wskazuje, że teza o „związku wstecz” daje się utrzymać tylko przy założeniu słabego pojęcia wartościowania, rozumianego jako równoważne z dowolnym rodzajem pragnień czy preferencji (w opozycji do silnego wartościowania, rozumianego jako ocena sprawcy, że działanie jest godne realizacji). Wówczas faktycznie, jeśli sprawca intencjonalnie α -uje, to wartościuje α -owanie

w tym znaczeniu, że ma jakieś pragnienie czy preferencje, aby α -ować (Pears 1982, s. 194–195). Te dwa rodzaje wartościowania odpowiadają używanym przez Melego pojęciom pragnienia w sensie motywacyjnym i ewaluacyjnym. Zatem zarówno Pears, jak i Mele wskazują przy tym, że działanie intencjonalne może zostać wywołane przez inne pragnienie, silniejsze w sensie słabego wartościowania (motywacji), lecz słabsze w sensie silnego (ewaluacji).

Także trzeci argument, oparty na odwrotnej tezie – głoszącej, że jeśli S żywi najlepszy sąd, by α -ować, to S najbardziej chce α -ować – Mele odpiera, wskazując ekwiwokację między dwoma znaczeniami pojęcia chcenia. Analizuje on wnioskowanie o następującej strukturze:

1. Jeśli S żywi w t najlepszy sąd, by α -ować, to S najbardziej chce α -ować w t .
2. Jeśli S najbardziej chce α -ować w t , to S intencjonalnie α -uje w t (o ile podejmuje jakiekolwiek działanie).

W: Nie jest możliwe, aby jednocześnie S żywił w t najlepszy sąd, by α -ować, oraz nie α -ował w t (o ile podejmuje jakiekolwiek działanie).

Przesłanka jest prawdziwa tylko po przyjęciu rozumienia ewaluacyjnego, zaś fałszywa na gruncie motywacyjnego, natomiast druga odwrotnie: prawdziwa w sensie motywacyjnym, ale fałszywa w ewaluacyjnym (Mele 1987, s. 31–49).

Tak przedstawia się argumentacja Melego przeciwko sceptykom. Warto dodać, że zdaniem tego autora za tezę o możliwości wystąpienia rozbieżności między motywacją a ewaluacją, uchwyconej w twierdzeniu, że „siła motywacyjna naszych pragnień nie zawsze odpowiada ewaluacji przedmiotów naszych pragnień” (Mele 2012, s. 119), a wyrażającej stanowisko eksternalizmu motywacyjnego (Stroud, Svirsky 2021; Milczarek 2022, s. 150), skądinąd przemawiają silne argumenty. Mele wskazuje trzy odmienne źródła świadectw.

Po pierwsze, są to „przykłady z rzeczywistego świata” (Mele 1987, s. 37): poddanie się chęci zapalenia papierosa czy lękowi przed lataniem samolotem, albo ulegnięcie pokusie wypoczywania w czasie przeznaczonym na pracę. Autor wskazuje, że już Sokrates miał twierdzić, że większość ludzi utrzymuje, iż niekiedy zdarza się im działać wbrew temu, o czym wiedzą, że jest najlepsze – co można potwierdzić ankietowo (Mele 2012, s. 87–89). W dwóch nieformalnych eksperymentach średnia odpowiedzi, zgodnie z którymi badanym studentom (kolejno 49 i 51 uczestników) zdarza się działać wbrew najlepszemu sądowi, wynosiła kolejno 1.224 i 1.9 w skali 1–7 (1 – zdecydowanie tak, 7 – zdecydowanie nie)¹¹.

¹¹ W drugim pytaniu dodano wprost, że działania te mają być wolne. Kontynuację badań nad słabością woli w nurcie filozofii eksperymentalnej można znaleźć m.in. w: May, Holton 2012; Beebe 2013; Doucet, Turri 2014; Sousa, Mauro 2015; Rosas i in. 2018; Messerli i in. 2022.

Po drugie, na rzecz tej samej tezy można sformułować argumenty z wykorzystaniem eksperymentów myślowych: na przykład wyobrazić sobie sytuację, w której sprawca wie, że padł ofiarą niegodziwego neuronaukowca, który implementuje bądź wzmacnia pewne pragnienia, a ich realizacja doprowadzi do przykrych konsekwencji. Wówczas ów sprawca prawdopodobnie będzie nisko oceniał przedmioty tych pragnień mimo ich dużej siły motywacyjnej (Mele 1987, s. 37–38). Innego przykładu obecności motywacji bez ewaluacji dostarcza „radioman” z artykułu Warrena Quinna – ktoś z dyspozycją do włączania jakiegokolwiek radioodbiornika w jego zasięgu, kto jednocześnie nie widzi żadnej wartości w tym, aby były one włączone (Quinn 1993). Alternatywnie możemy pomyśleć o istnieniu urządzeń pozwalających na dowolne manipulowanie siłą motywacyjną pragnień – na przykład osłabianie siły oddziaływania niektórych pragnień na życzenie – które nie wpływałyby na proces ewaluacji (Mele 1987, s. 28).

Po trzecie, za niezgodnościami między siłą motywacyjną a ewaluacją mają przemawiać dane empiryczne. Liczne badania (m.in. George’a Ainsle’ego oraz Waltera Mischela) świadczą na rzecz tezy, że siła motywacyjna rośnie w miarę zbliżania się gratyfikacji, a także iż daje się ona modyfikować poprzez różne strategie poznawcze – przede wszystkim nakierowanie uwagi na określone (pozytywne lub negatywne) aspekty pożądanego stanu rzeczy – oraz przez zastosowanie innych technik samokontroli, np. formułowania osiągalnych i ściśle wyznaczonych osobistych postanowień (Mele 1987, s. 84–91; Mele 2012, s. 77–82, 96–104)¹².

Ponadto do wprowadzenia takiego rozróżnienia mogą skłaniać obserwacje z pól ontogenezy (rozwoju osobniczego) i filogenezy (rozwoju gatunkowego). Intuicyjne wydaje się twierdzenie (choć oczywiście podlega ono częściowo empirycznemu sprawdzeniu), że istnieje odstęp między okresem, w którym rozwijają się kompetencje odpowiadające za motywację, od tych odpowiadających za ewaluację – a więc że istnieje moment, w którym mamy prawo uznać, że człowiek jest już zmotywowany do podejmowania określonych zachowań, ale nie jest w stanie oceniać ich rezultatów (ze względu na jeszcze nierozwinięte kompetencje o charakterze kognitywnym). Ta druga zdolność może być skorelowana z rozwojem kompetencji samokontroli w ok. 2–3 roku życia (Vaughn i in. 1984). Analogicznie można spekulować, że dopiero w pewnym momencie swojej historii ewolucyjnej ludzie (lub ich przodkowie) zaczęli oceniać przedmioty swoich pragnień, chociaż już wcześniej byli zmotywowani do podejmowania różnego rodzaju zachowań.

¹² Aktualne omówienie empirycznych podstaw dla tej tezy można znaleźć w: Heinzelmann 2023.

3. Druga ekwiwokacja: pojęcie racjonalności

Przeciwko koncepcji Melego można argumentować na wiele sposobów – jeden z nich, któremu poświęcona jest dalsza część pracy, dotyczy normatywnego wymiaru działań akratycznych. Pojęcie normatywności odgrywa kluczową rolę w dyskusji nad słabością woli, gdyż zasadniczym pytaniem, na jakie powinna odpowiadać adekwatna teoria akrazji, jest to, dlaczego intuicyjnie uznajemy działania wbrew najlepszemu sądowi za irracjonalne i o jakiej nieracjonalności jest tu właściwie mowa¹³.

Na problem z normatywnym wymiarem przypisać słabości woli wskazywał już Donald Davidson:

Paradoks leżący u podstaw irracjonalności, którego żadna teoria nie może całkiem uniknąć, jest taki: jeśli wyjaśnimy ją zbyt dobrze, zmienimy ją w zawołowaną formę racjonalności; natomiast jeśli przypisujemy niespójność zbyt swobodnie, jedynie upośledzamy naszą zdolność rozpoznania irracjonalności, wyzbywając się tła racjonalności niezbędnego do uzasadnienia jakiegokolwiek rozpoznania (Davidson 1982, s. 303).

Rozwiązanie tego paradoksu wymaga zatem sformułowania wyjaśnienia słabości woli, które jednocześnie będzie uwzględniało dwa aspekty: (a) lokalną irracjonalność działania akratycznego, (b) ogólnie racjonalistyczny charakter pojęcia działania, wynikający z umiejscowienia działań w porządku racji. Jednak, jak będę argumentował w dalszej części tekstu, propozycja Melego nie jest w stanie dostarczyć satysfakcjonującego uzasadnienia dla przypisania irracjonalności sprawcy działającemu akratycznie, a pozory obecności takiego ugruntowania wynikają z ekwiwokacji w stosowaniu pojęcia racjonalności.

Zdaniem Melego spełnienie drugiego z tych wymogów jest możliwe, gdyż – jak zostało wskazane powyżej – działanie akratyczne stanowi działanie ze względu na pewne racje i w tym ograniczonym sensie jest ono minimalnie racjonalne. Pozostaje więc kwestia wyjaśnienia, w jakim sensie działania akratyczne są irracjonalne. Naturalnym wyborem jest tu pojęcie racjonalności rozumianej jako wewnętrzna niespójność sprawcy. Ta odnosi się bowiem do wymogów takich jak: nieżywienie sprzecznych przekonań, nieformułowanie wykluczających się intencji, niepodjęmowanie intencji do działań, o których jesteśmy przekonani, że nie będziemy w stanie ich wykonać, a także zasady powściągliwości – działania zgodnie z najlepszym sądem (Żuradzki 2014, s. 36–37)¹⁴. Ostatni z wyróżnionych typów stanowi właśnie wymóg dotyczący

¹³ Terminy „nieracjonalność” i „irracjonalność” używane są w tej pracy zamiennie.

¹⁴ Na gruncie rozróżnienia trzech poziomów racjonalności, zaproponowanych w innym kontekście przez Josepha Agassiego i Iana Jarvie’ego, racjonalność wymagana dla racjonalizacji odpowiadałaby racjonalności₁ – nakierowaniu działania na cel, natomiast powściągliwość działań – racjonalności₃, właściwej dla działania zgodnego z najlepszym dostępnym zestawem reguł

spójności nastawień sądzeniowych z działaniem, który jest naruszany przez sprawcę działającego akratycznie. Zatem także i tu należy wskazać na pewną ekwiwokację. Pojęcie racjonalności jest dwuznaczne, gdyż dotyczy albo racjonalności minimalnej (racjonalności_m), albo racjonalności jako spójności (racjonalności_s). W końcu, po uznaniu, że działania akratyczne są racjonalne_m, należy zapytać, co w koncepcji Melego pozwala na uznanie działań akratycznych za nieracjonalne_s.

W odpowiedzi na to pytanie można wskazać na dwa – niestety dość lakonicznie scharakteryzowane przez Melego – elementy, które odróżniają słabość woli od typowych (racjonalnych_s) działań sprawcy. Po pierwsze, jest to normatywne zobowiązanie płynące z najlepszego sądu. Zdaniem tego autora jest on źródłem zobowiązania ewaluacyjnego, a jego normatywna siła bierze się stąd, że jest niepełnym wnioskiem deliberacji praktycznej¹⁵: rozstrzyga pytanie, co (z perspektywy sprawcy) najlepiej byłoby zrobić (Mele 1995, s. 71), a jego podstawową funkcją jest kierowanie postępowaniem (tamże, s. 25). Mele wspomina normatywny wymiar teorii decyzji: racjonalny sprawca powinien realizować swoje preferencje (tamże, s. 76–77). Sąd ewaluacyjny zobowiązuje do działania w tym sensie, że działanie wbrew niemu jest subiektywnie irracjonalne, ponieważ sprawca, działając akratycznie, nie ma (ze swojego punktu widzenia) lepszych racji na rzecz działania innego niż to, o którym mówi treść najlepszego sądu, a mimo to podejmuje właśnie takie działanie (Mele 1987, s. 5–6; Mele 1995, s. 71).

Po drugie więc, istotna dla przypisania nieracjonalności_s jest pierwszoosobowa perspektywa samych sprawców, którzy – jeśli działają akratycznie – ze swojego pierwszoosobowego punktu widzenia postrzegają sami siebie jako wewnętrznie niespójnych. Podczas analizy przypadków słabości woli – potocznych, wyobrażonych czy uzyskanych w warunkach eksperymentalnych (por. wyżej § 2) – Mele formułuje interpretacje z uwzględnieniem perspektywy samych akratycznych sprawców, co nie dziwi, jeśli pod uwagę weźmie się fakt, że określa on najlepsze sądy jako „pierwszoosobowe sądy praktyczne” (Mele 1987, s. 5), a w definicji działań ściśle niepowściągliwych dodaje, że sprawca „świadomie utrzymuje sąd” (s. 7), wbrew któremu działa. Ze względu na te uwagi uprawnione jest uznanie, że dla Melego uwzględnienie pierwszo-

(Agassi, Jarvie 1979, s. 432). Przy czym w teorii działania byłyby to reguły – normy racjonalności – rozumiane internalistycznie, bo wyznaczone przez wewnętrzną spójność między stanami mentalnymi oraz działaniami sprawcy.

Dodajmy, że o ile zachowania, które nie są racjonalne w pierwszym z tych sensów – racjonalności jako racjonalizowalności – raczej w ogóle nie powinny zostać uznane za działania, to działania naruszające spójność sprawcy – irracjonalne w drugim znaczeniu – wciąż nimi pozostają.

¹⁵ Niepełnym, ponieważ Mele zgadza się ze standardowym poglądem, że ostatecznie to intencja stanowi efekt deliberacji (Mele 1992, s. 138–139).

osobowego, świadomego rozpoznania niespójności sprawcy jest warunkiem koniecznym irracjonalności_s.

Wyłaniająca się z prac autora *Irrationality* odpowiedź na pytanie o ugruntowanie zobowiązania ewaluacyjnego jest pozornie nieskomplikowana: sprawca powinien działać powściągliwie, ponieważ on sam ocenia to działanie jako lepsze – mowa jest tu w końcu o jego własnej ewaluacji i jego wewnętrznej spójności. Jednak wydaje się, że o ile racjonalność_m daje się wpisać w ogólny filozoficzny projekt Melego, to nieracjonalność_s dodana zostaje *ad hoc*. Jest tak, ponieważ autor jest zwolennikiem szeroko pojętego (por. Heil, Mele 1991) funkcjonalizmu w kwestii natury umysłu. Charakteryzuje on bowiem stany i procesy mentalne (w szczególności: najlepsze sądy, intencje i samokontrolę) poprzez wskazanie ich związków przyczynowych z informacyjnym „wejściem”, innymi stanami (lub procesami) mentalnymi oraz działaniowym „wyjściem” (Mele 1987, 1991, 1992).

O ile z interpretacji Scotta Sehona wynika, że Mele jest zwolennikiem pełnej redukcji funkcjonalnej – uwzględniającej kontrowersyjne założenie, że po scharakteryzowaniu własności funkcjonalnych uda się utożsamić je z pewnymi własnościami postulowanymi na gruncie nauk szczegółowych (Sehon 1997, s. 198–204)¹⁶ – to dla aktualnych rozważań takie założenie nie jest konieczne. Możemy zatrzymać się na pierwszym z wyróżnionych przez Neda Blocka etapów redukcji, czyli charakterystyce stanów mentalnych na poziomie sieci systematycznych, acz nie bezwyjątkowych relacji przyczynowych (Block 2011).

Jednym z zasadniczych argumentów przeciwko funkcjonalistycznemu podejściu do przypisywania stanów mentalnych jest niezdolność do uchwycenia norm racjonalności dla samych działań (McDowell 1985). W niektórych odmianach funkcjonalizmu sugeruje się zatem, że normy racjonalności w dużej mierze odpowiadają relacjom przyczynowym (Fodor 2001) lub – w ramach tzw. normatywnego funkcjonalizmu – uznaje się, że stany funkcjonalne są istotowo normatywne, a więc nie tyle wywołują, co powinny wywoływać daną reakcję behawioralną (Zangwill 1998). Jednak w przypadku Melego problem ten w ogóle nie zostaje podniesiony. Propozycja autora ogranicza się do uwzględniania przyczynowego opisu racjonalności_m – relacji kauzalnych między działaniami a racjami (na wzór klasycznej przyczynowej teorii działania intencjonalnego) – lecz nie zawiera próby wpisania racjonalności_s w założony funkcjonalistyczny model umysłu.

¹⁶ Jest ono kontrowersyjne w tym kontekście, gdy weźmie się pod uwagę twierdzenie Melego, że „niezależnie od tego, jak zaawansowana będzie neuronauka, tak długo jak będziemy przypisywać sprawcom moralną odpowiedzialność za poszczególne czyny, będziemy potrzebować sposobu odróżnienia działań intencjonalnych od nieintencjonalnych, a to będzie wymagało wykorzystania pojęć intencjonalistycznych” (Mele 1992, s. 42).

W konsekwencji Mele może starać się wyjaśnić irracjonalność_s akrazji, lecz jest to wyjaśnienie co najwyżej niepełne. Jest tak, bowiem zobowiązanie płynące z najlepszych sądów i wymagające świadomego samorozumienia się sprawcy jako spójnego podmiotu nie mają teoretycznych powiązań z ogólną koncepcją Melego. Nie jest też wcale oczywiste, że w ogóle dają się one w nią wpisać, gdyż przy próbie powiązania obu tych porządków mogą pojawić się nieusuwalne sprzeczności. Aby ocenić, czy tak właśnie jest, warto przyjrzeć się dokładniej, w jaki sposób taka redukcja mogłaby zostać przeprowadzona i jakie trudności się z tym wiążą.

W pierwszej kolejności rozważmy kwestie zobowiązań ewaluacyjnych, których źródłem mają być najlepsze sądy. Podstawową trudnością w funkcjonalizacji najlepszych sądów jako nastawień racjonalizujących_s jest napięcie między dwoma *prima facie* przekonującymi tezami:

- a) Najlepszy sąd jest źródłem zobowiązania, ponieważ wyraża ewaluację przedmiotów pragnień sprawcy.
- b) Istnieje zasadnicza trudność z przeprowadzeniem redukcji funkcjonalnej zdolności ewaluacji.

O ile pierwszej z tez nie trzeba dowodzić, gdyż oczywiste jest, że to nie komponent motywacyjny racji odpowiada za zobowiązania sprawcy (*vide* scenariusze zniewalających pragnień), to druga wymaga objaśnienia. Gdyby głosiła ona, że całkiem nie da się przeprowadzić redukcji funkcjonalnej ewaluacji, to łącznie z (a) przesądzałaby o porażce projektu Melego. Zatem konieczne jest znalezienie takiej interpretacji zdolności ewaluacji, która poddaje się opisowi w siatce relacji przyczynowych.

To, czy taka interpretacja będzie dostępna, wydaje się zależeć od przyjętego rozumienia, czym jest ewaluacyjny komponent racji (czyli ewaluacja przedmiotu danego pragnienia). Aby odpowiedzieć, dlaczego miałyby tak być, w pierwszej kolejności założmy, że to racje same w sobie posiadają zarówno komponent motywacyjny, jak i ewaluacyjny (a tak zdaje się sugerować Mele). W takim wypadku jednak tym, co odpowiada za przyczynowe oddziaływanie racji, byłby wciąż tylko ich komponent motywacyjny, ponieważ to od siły motywacyjnej zależy, w jakie relacje przyczynowe będzie wchodzić dane nastawienie lub – gdy mowa o racjach – para nastawień¹⁷. Wówczas komponent ewaluacyjny okazuje się przyczynowo nierelevantny – a jeśli nie jest istotny przyczynowo, nie może podlegać redukcji funkcjonalnej.

W przeciwnym wypadku – przy uznaniu, że komponent ewaluacyjny nie jest neutralny przyczynowo – dochodziłoby do regresu. Sprawca mógłby bo-

¹⁷ Warto tutaj podkreślić, że przekonanie będące elementem racji za α -owaniem nie może wyrażać pozytywnej oceny efektu α -owania, gdyż jest to jedynie przekonanie instrumentalne.

wiem działać ze względu na rację, którą ocenia jako gorszą (ewaluacyjnie słabszą), nie dlatego, że jej komponent motywacyjny jest silniejszy, ale z powodu mocniejszego oddziaływania przyczynowego jej komponentu ewaluacyjnego. Aby wyjaśnić taki przypadek akrazji, konieczne byłoby ponowne rozróżnienie motywacji i ewaluacji – tym razem dla samego komponentu ewaluacyjnego (co oczywiście można powtórzyć w odniesieniu do kolejnych komponentów ewaluacyjnych „niższego rzędu”).

Istnieje możliwość uniknięcia tego problemu poprzez wskazanie alternatywnej interpretacji wyrażenia „komponent ewaluacyjny”, ale wymaga to poniesienia pewnych kosztów. Mianowicie można uznać, że ewaluacja jest zewnętrzna wobec racji, a więc że proces deliberacji nie polega wyłącznie na ważeniu samych racji, lecz potrzebne jest jeszcze coś – gdzie „tym czymś” byłaby właśnie ewaluacja – aby doprowadzić do działania. Wyzwaniem dla tego rozwiązania jest także problem regresu (z tym że dotyczący czego innego niż ten w akapicie powyżej): taka ewaluacja nie może być prowadzona ze względu na jakiekolwiek racje, gdyż jej przedmiotem są właśnie racje – inaczej konieczne byłoby także wyjaśnienie roli tych racji, ze względu na które prowadzona jest ewaluacja racji, do czego należałoby zapostulować ewaluację na kolejnym poziomie (Railton 2009). Rozwiązaniem Petera Railtona, ale także takich teoretyków jak Antonio Damasio czy Jonathan Haidt, jest uznanie, że ewaluacja jest procesem raczej afektywnym niż intelektualnym – opartym na często nieuświadomionych przeczuciach, a nie na świadomych (czy nawet uświadomialnych) stanach mentalnych (Damasio 2011; Haidt 2001)¹⁸.

Jest to rewizyjna koncepcja ewaluacji, która może wzbudzać następujące wątpliwości. Po pewnym dostrojeniu stosuje się wobec niej Jaegwona Kima krytyka „epistemologii znaturalizowanej” (Kim 1988/2016): mianowicie można uznać, że ewaluacja w omawianym sensie nie może zagwarantować uzasadnienia dla działania (z perspektywy sprawcy) ani nie pozwala go wyjaśnić (z trzecioosobowego punktu widzenia). Mowa tu o uzasadnieniu i wyjaśnieniu porównawczym, czyli odpowiedzi na pytanie, dlaczego sprawca podjął pewne działanie zamiast innego – na które zwolennicy emocjonalnej koncepcji ewaluacji nie mogą odpowiedzieć, powołując się na normatywną siłę racji, a jedynie na przyczynę w postaci afektywnego „popchnięcia” w stronę realizacji tej lub innej racji. Pojawia się także pytanie, czy nie jest to wyłącznie rozwiązanie problemu w sensie angielskiego terminu *dissolution* – zignorowanie go z poziomu osobowej sprawczości poprzez sprowadzanie do poziomu subosobowego¹⁹.

Nie rozstrzygając, czy takie rozumienie ewaluacji byłoby dla Melego rozwiązaniem do przyjęcia, rozważmy drugi z elementów mających świadczyć

¹⁸ Również Karen Jones przedstawia umiarkowaną wersję tego stanowiska (Jones 2003).

¹⁹ Cały ten akapit jest wynikiem przetworzenia myśli, które oryginalnie sformułował dr Michał Barcz.

o irracjonalności, słabej woli – pierwszoosobowym rozpoznaniu niespójności objawiającej się w akratycznym postępowaniu sprawcy. Sam fakt, że Mele powołuje się na świadomość sprawcy, nie jest jeszcze zarzutem wobec jego stanowiska, gdyż mniej lub bardziej eksplicytna praktyka powoływania się na warunek świadomości w rozważaniach dotyczących norm racjonalności jest powszechna w filozofii działania – wyklucza się np. jedynie świadome żywienie przekonań niepopartych żadnymi świadectwami (Winters 1979, s. 245) lub świadome żywienie przekonań sprzecznych (Priest 1986, s. 103)²⁰.

Mimo to warto zauważyć, że czym innym jest pytanie o dostęp sprawcy do własnych stanów mentalnych niż pytanie o ten dostęp na gruncie funkcjonalizmu. O ile były czynione próby poradzenia sobie z wpisaniem świadomości własnego umysłu w ten projekt filozoficzny, to istnieją mocne argumenty przeciwko redukcji funkcjonalnej pierwszoosobowego dostępu (Bealer 1997)²¹. Jednak tym, co może niepokoić w ramach koncepcji Melego nawet bardziej od konkretnych argumentów wymierzonych w możliwość redukcji funkcjonalnej świadomości, jest bezkrytycznie założony potoczny obraz sprawcy jako świadomego swojej wewnętrznej spójności podmiotu.

O ile charakterystyka funkcjonalna stanowi przykład tzw. wyjaśnienia wertykalnego, czyli wyjaśnienia dekompozycyjnego – wskazania i opisanie części systemu umożliwiających realizację zdolności systemowych (Bermudez 2000, s. 63–82; Gładziejewski 2012, s. 46) – to Melego rozumienie sprawcy, konieczne do uznania działań akratycznych za irracjonalne, przypomina raczej przedteoretyczne rozumienie osoby dostępne w ramach „manifestującego się” (nazywanego też „jawnym” czy „potocznym”) obrazu świata (Sellars 1963/1991). Zdaniem Wilfrida Sellarsa na to, czym z tej perspektywy jest osoba, składają się dwa główne elementy: (a) „określa ona sama siebie jako cechującą się jednością”, (b) „jej tożsamość powstaje dzięki samorozumieniu i samoocenie” (Bremer 2012, s. 32). Obydwa te elementy są obecne w omawianej koncepcji racjonalności, gdzie w (a*) indywidualnej perspektywie wewnętrznie spójnego sprawcy (b*) konstituują się normy racjonalności, mające gwarantować, by jego własne działanie było dla niego zrozumiałe – zharmonizowane z odpowiednimi stanami mentalnymi.

Tymczasem projekt naturalizacji irracjonalności (w ramach którego funkcjonalizm jest jedną z opcji) może wymagać odejścia od takiego rozumienia sprawcy (Antony 2000). W szczególności pojawia się tu pytanie, dlaczego na gruncie naturalistycznie rozumianej racjonalności należałoby przyznać prymat

²⁰ Oczywiście pojawia się pytanie, który z rozlicznych sensów pojęcia świadomości jest tu istotny. Jest to sprawa otwarta – na potrzeby dyskusji założone zostaje jedynie, że jest to świadomość konieczna do zagwarantowania pierwszoosobowego rozpoznania własnej niespójności przez sprawcę.

²¹ Por. odpowiedź na argument George’a Bealera w: Båve 2017.

niespójnościom, które są uświadamiane sobie przez sprawcę. Wydaje się, że równie istotne dla realizacji celów sprawcy są naruszenia związków przyczynowych między różnymi nastawieniami sądzeniowymi oraz nastawieniami a działaniem, które nie są uświadomione; podobnie nic nie wyklucza istnienia mechanizmów rozpoznających i wdrażających racje w działanie bez udziału świadomości.

Powyższe uwagi prowadzą więc do trzech wniosków. Po pierwsze, racjonalność w sensie racjonalizowalności jest czymś zasadniczo różnym od racjonalności w sensie spójności. Po drugie, obydwa te sensy są istotne w dyskusji o słabości woli. Wreszcie po trzecie, dopóki w koncepcji Melego nie zostanie uwzględniona racjonalność w drugim z tych sensów, dopóty – mimo wszystkich swoich zalet – nie będzie to satysfakcjonująca koncepcja działań akratycznych.

Podsumowanie

Jak zauważył sam Mele, problem słabości woli rodzi się na przecięciu czysto intelektualnego oraz czysto przyczynowego spojrzenia na ludzkie działanie. Z punktu widzenia pierwszego nie istnieją żadne działania akratyczne, gdyż każde działanie jest całkowicie racjonalne, natomiast z perspektywy drugiego niedostrzegalny jest problem działań akratycznych, ponieważ człowiek postępujący niepowściągliwie jest motywacyjnie nieodróżnialny od swojego powściągliwego odpowiednika (Mele 2012, s. 60–61). Zmaganie się z uwzględnieniem, a najlepiej – pogodzeniem obu tych perspektyw wydaje się szczególnym przypadkiem bardziej ogólnych trudności, z którymi mierzy się naturalistycznie zorientowana filozofia, a których źródeł Sellars upatrywał w napięciu między naukowym a manifestującym się obrazem świata.

Niniejszy tekst stanowił omówienie właśnie takich zmagających. W pierwszej kolejności zrekonstruowana została Melego koncepcja działań akratycznych, ze szczególnym wskazaniem, dlaczego przejawy słabości woli mogą być uznane za działania intencjonalne oraz wolne. Następnie zrealizowany został podstawowy cel artykułu – na tle koncepcji Melego wskazałem i krytycznie omówiłem dwie ekwiwokacje w dyskusji o słabości woli. Jak starałem się wykazać, pierwsza z nich świadczy o nietrafności argumentów sceptycznych wobec możliwości działania wbrew najlepszemu sądowi, natomiast druga zmusza do uznania koncepcji Melego za nieadekwatną, bo niewystarczającą do wyjaśnienia irracjonalności działań akratycznych.

Bibliografia

- Agassi A., Jarvie I.C. (1979), *The Rationality of Dogmatism*, w: ciż (red.), *Rationality: The Critical View*, Dordrecht: Springer, s. 431–443.
- Anscombe G.E.M. (1957), *Intention*, Oxford: Basil Blackwell.
- Antony L.M. (2000), *Naturalised Epistemology, Morality, and the Real World*, „Canadian Journal of Philosophy” 30 (supp. vol. 26), s. 103–137.
- Båve A. (2017), *Self-Consciousness and Reductive Functionalism*, „Philosophical Quarterly” 67 (266), s. 1–21.
- Bealer G. (1997), *Self-Consciousness*, „The Philosophical Review” 106, s. 69–117.
- Beebe J.R. (2013), *Weakness of Will, Reasonability, and Compulsion*, „Synthese” 190 (18), s. 4077–4093.
- Bermudez J.L. (2000), *Personal and Sub-Personal: A Difference Without a Distinction*, „Philosophical Explorations” 3 (1), s. 63–82.
- Bigelow J., Dodds S.M., Pargetter R. (1990), *Temptation and the Will*, „American Philosophical Quarterly” 27 (1), s. 39–49.
- Block N. (2011), *Functional Reduction*, w: M. Sabatés, D. Sosa, T. Horgan (red.), *Supervenience in Mind: A Festschrift for Jaegwon Kim*, Cambridge, MA: MIT Press.
- Bremer J. (2012), *Dwa obrazy świata: manifestujący się i naukowy*, „Roczniki Filozoficzne” 60 (1), s. 27–49.
- Burch M. (2018), *Making Sense of Akrasia*, „Phenomenology and the Cognitive Sciences” 17 (5), s. 939–971.
- Damasio A. (2011), *Błąd Kartezjusza: emocje, rozum i ludzki mózg*, przeł. M. Karpiński, Poznań: Dom Wydawniczy Rebis.
- Davidson D. (1963/1980), *Actions, Reasons, and Causes*, w: tenże, *Essays on Actions and Events*, Oxford: Clarendon Press 1980, s. 3–21.
- Davidson D. (1970/1997), *Jak jest możliwa słabość woli?*, przeł. W.J. Popowski, w: J. Hołówka (red.), *Filozofia moralności. Postanowienie i odpowiedzialność moralna*, Warszawa: Wydawnictwo Spacja – Fundacja Aletheia 1997, s. 81–106.
- Davidson D. (1982), *Paradoxes of Irrationality*, w: R. Wolheim, J. Hopkins (red.), *Philosophical Essays on Freud*, Cambridge: Cambridge University Press, s. 289–305.
- Doucet M., Turri J. (2014), *Non-Psychological Weakness of Will: Self-Control, Stereotypes, and Consequences*, „Synthese” 191 (16), s. 3935–3954.
- Fodor J. (2001), *Eksperci od wiązków: język myślowy i jego semantyka. Wykłady im. Jeana Nicod z roku 1993*, przeł. M. Gokieli, Warszawa: Fundacja Aletheia.
- Gładziejewski P. (2012), *William Ramsey o psychologii potocznej, racjonalności i pojęciu reprezentacji w naukach kognitywnych*, „Diametros” 31, s. 33–55.
- Heil J., Mele A. (1991), *Mental Causes*, „American Philosophical Quarterly” 28 (1), s. 61–71.
- Heinzelmann N. (2023), *Weakness of Will and Delay Discounting*, Oxford: Oxford University Press.
- Holton R. (1999), *Intention and Weakness of Will*, „The Journal of Philosophy” 96 (5), s. 241–262.

- Holton R. (2003), *How Is Strength of Will Possible?*, w: C. Tappolet, S. Stroud (red.), *Weakness of Will and Practical Irrationality*, Oxford: Clarendon Press, s. 39–67.
- Jackson F. (1984), *Weakness of Will*, „Mind” 93 (369), s. 1–18.
- Janikowski W. (2016), *Wolna wola i użyteczność. Podstawy etyki utilitaryzmu preferencji*, Warszawa: Wydawnictwo Naukowe Semper.
- Jeffrey R.C. (1974), *Preference Among Preferences*, „The Journal of Philosophy” 71 (13), s. 377–391.
- Johnson C.M. (2021), *Resolutions, Salient Reasons, and Weakness of Will*, „Synthese” 198, s. 5115–5138.
- Jones K. (2003), *Emotion, Weakness of Will, and the Normative Conception of Agency*, w: A. Hatzimoysis (red.), *Philosophy and the Emotions*, Cambridge: Cambridge University Press, s. 181–200.
- Kim J. (1988/2016), *What is „Naturalized Epistemology”?*, w: J.E. Tomberlin (red.), *Philosophical Perspectives* 2, Atascadero, CA: Ridgeview Publishing, s. 381–406; tłum. pol. *Czym jest „epistemologia znaturalizowana”?*, przeł. M. Garbowski, „Roczniki Filozoficzne” 64 (3), s. 115–141.
- Lazar A. (1996), *Akrazja a zasada powściągliwości lub jeszcze raz o tym, co Achilles powiedziałby żółwiowi*, w: U.M. Żegleń (red.), *Dyskusje z Donaldem Davidsonem o prawdzie, języku i umyśle*, Lublin: Towarzystwo Naukowe KUL, s. 309–332.
- Machciewicz-Grad A. (2020), *Akrazja jako dysfunkcja woli. Krytyczna analiza Davidsonowskiej koncepcji słabej woli oraz zarys alternatywnego rozwiązania*, „Przegląd Filozoficzny – Nowa Seria” 1 (113) s. 211–233.
- May J., Holton R. (2012), *What in the World is Weakness of Will?*, „Philosophical Studies” 157 (3), s. 341–360.
- McDowell J. (1985), *Functionalism and Anomalous Monism*, w: E. LePore, B. McLaughlin (red.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell Publishers. s. 387–398.
- Mele A. (1987), *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*, New York: Oxford University Press.
- Mele A. (1989), *Motivational Internalism: The Powers and Limits of Practical Reasoning*, „Philosophia” 19 (4), s. 417–436.
- Mele A. (1991), *Akratic Action and the Practical Role of Better Judgement*, „Pacific Philosophical Quarterly” 71 (1), s. 33–47.
- Mele A. (1992), *Springs of Action*, New York: Oxford University Press.
- Mele A. (1995), *Autonomous Agents: From Self-Control to Autonomy*, New York: Oxford University Press.
- Mele A. (2003), *Motivation and Agency*, New York: Oxford University Press.
- Mele A. (2010), *Weakness of Will and Akrasia*, „Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition” 150 (3), s. 391–404.
- Mele A. (2012), *Backsliding: Understanding Weakness of Will*, New York: Oxford University Press.
- Mele A. (2014), *Self-Control, Motivational Strength, and Exposure Therapy*, „Philosophical Studies” 170, s. 359–375.
- Messerli M., Fink J., Reuter K. (2022), *The Varying Rationality of Weakness of the Will: an Empirical Investigation and its Challenges for a Unified Theory of Rationality*, „Synthese” 200 (365), s. 1–23.

- Michael M.T. (2019), *Passionate Akrasia*, „Philosophia” 47, s. 569–585.
- Milczarek K. (2022), *Słaba wola – w stronę eksternalizmu motywacyjnego*, „Edukacja Filozoficzna” 73, s. 145–162.
- Pears D. (1986), *Motivated Irrationality*, Oxford: Clarendon Press.
- Pears D.F., Pugmire D. (1982), *Motivated Irrationality*, „Proceedings of the Aristotelian Society” 56, s. 179–196.
- Priest G. (1986), *Contradiction, Belief and Rationality*, „Proceedings of the Aristotelian Society. New Series” 86, s. 99–116.
- Quinn W. (1993), *Putting Rationality in its Place*, w: tenże, *Morality and Action*. Cambridge: Cambridge University Press, s. 228–255.
- Rosas A., Bermudez J.P., Gutiérrez Cabrera J.A. (2018), *Is a Bad Will a Weak Will? Cognitive Dispositions Modulate Folk Attributions of Weakness of Will*, „Philosophical Explorations” 21 (3), s. 350–363.
- Sehon S. R. (1997), *Deviant Causal Chains and the Irreducibility of Teleological Explanation*, „Pacific Philosophical Quarterly” 78 (2), s. 195–213.
- Sellars W. (1963/1991), *Philosophy and the Scientific Image of Man*, w: tenże, *Science, Perception and Reality*, Atascadero: Ridgeview Publishing Company 1991; tłum. pol. *Filozofia a naukowy obraz człowieka*, w: tenże, *Nauka, percepcja, rzeczywistość*, przeł. S. Wilczewska, Warszawa: Wydawnictwo Naukowe PWN 2023, s. 1–58.
- Sousa P., Mauro C. (2015), *The Evaluative Nature of the Folk Concepts of Weakness and Strength of Will*, „Philosophical Psychology” 28 (4), s. 487–509.
- Sripada C.S. (2014), *How is Willpower Possible? The Puzzle of Synchronic Self-Control and the Divided Mind*, „Noûs” 48 (1), s. 41–74.
- Stroud S., Svirsky L. (2021), *Weakness of Will*, w: E.N. Zalta (red.), *The Stanford Encyclopedia of Philosophy*, Winter Edition, <https://plato.stanford.edu/archives/win2021/entries/weakness-will/>
- Tenenbaum S. (1999), *The Judgement of a Weak Will*, „Philosophy and Phenomenological Research” 59 (4), s. 875–911.
- Vaughn B.E., Kopp C.B., Krakow J.B. (1984), *The Emergence and Consolidation of Self-Control from Eighteen to Thirty Months of Age: Normative Trends and Individual Differences*, „Child Development” 55 (3), s. 990–1004.
- Walker A.F. (1989), *The Problem of Weakness of Will*, „Noûs” 23 (5), s. 653–676.
- Watson G. (1977), *Skepticism about Weakness of Will*, „The Philosophical Review” 86 (3), s. 316–339.
- Winters, B. (1979), *Believing at Will*, „Journal of Philosophy” 76 (5), s. 243–256.
- Zangwill N. (1998), *Direction of Fit and Normative Functionalism*, „Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition” 91 (2), s. 173–203.
- Żuradzki T. (2014), *Trzeci rodzaj normatywności – wymóg reagowania na racje*, „Filozofia Nauki” 22 (4), s. 35–51.

K a r o l M i l c z a r e k

**Alfred Mele: Two Equivocations in the Dispute
over the Weakness of Will****Keywords:** *akrasia, desire, A. Mele, rationality, weakness of will*

The paper provides a commentary on two equivocations relevant to Alfred Mele's argumentation in the dispute over the problem of weakness of will. The first one, concerning the notion of desire (used in either an evaluative or motivational sense), allows Mele to demonstrate the invalidity of the objections against the possibility of weak-willed action. The second, referring to the notion of rationality (understood either as explainability in intentional categories or as the internal consistency of the agent), is found in Mele's own argumentation. As a consequence, Mele is able to counter the arguments against the possibility of weak-willed action but does not provide sufficient grounds for recognizing akratic actions as irrational.