

BERNARD WILLIAMS

### RACJE WEWNĘTRZNE I ZEWNĘTRZNE\*

Zdania o postaci „*P* posiada rację, żeby  $\alpha$ -ować” lub „Istnieje racja dla *P*, żeby  $\alpha$ -ować” (gdzie  $\alpha$  zastępuje jakiś czasownik oznaczający działanie) da się na pierwszy rzut oka interpretować na dwa sposoby. Zgodnie z pierwszą interpretacją prawdziwość takiego zdania implikuje — z grubsza rzecz biorąc — że *P* ma jakiś motyw, który zostanie urzeczywistniony lub wprowadzony w życie przez  $\alpha$ -owanie. Jeśli natomiast okaże się, że tak nie jest, to zdanie takie jest fałszywe: istnieje pewien odnoszący się do celów podmiotu warunek, który gdy nie zostanie spełniony, to — zgodnie z tą interpretacją — powiedzenie, że *P* posiada rację, żeby  $\alpha$ -ować, nie będzie prawdziwe. Zgodnie z drugą interpretacją nie istnieje żaden taki warunek, a zdanie na temat racji nie zostanie sfalsyfikowane przez brak odpowiedniego motywu. Pierwszą interpretację nazwę „wewnętrzną”, a drugą — „zewnątrzną”. (Biorąc pod uwagę te dwie interpretacje i dwie podane powyżej formy zdań, można zasadnie przypuszczać, że pierwszemu zdaniu w bardziej naturalny sposób przypisuje się interpretację wewnętrzną, a drugiemu — zewnętrzną, ale to nie znaczy, że każde z tych sformułowań poddaje się tylko jednej z interpretacji).

---

Prof. BERNARD WILLIAMS (1929–2003) — brytyjski etyk i filozof, wykładał m.in. na uniwersytetach w Cambridge, Oksfordzie, Kalifornijskim w Berkeley. Autor m.in.: *Problems of the Self* (1973), *Utilitarianism: For and Against* (1973 – wraz z J.J.C. Smartem), *Ethics and The Limits of Philosophy* (1985), *Shame and Necessity* (1993). W języku polskim ukazały się: *Ile wolności powinna mieć wola* (1999), *Moralność. Wprowadzenie do etyki* (2001).

\* Tekst pierwotnie ukazał się w książce: Ross HARRISON (red.), *Rational Action* (Cambridge: Cambridge University Press, 1980), 17–28. Podstawa tłumaczenia: Bernard WILLIAMS, „Internal and External Reasons”, w: *Moral Luck: Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981), s. 101–113. © Cambridge University Press 1979, translated with permission.

Tłumaczenie oraz zakup licencji pozwalającej na publikację polskiego tłumaczenia zostały sfinansowane z programu Ministra Nauki i Szkolnictwa Wyższego pod nazwą „Narodowy Program Rozwoju Humanistyki” (nr 0068/NPRH4/H2b/83/2016).

Jedynie dla wygody będę odnosił się niekiedy do „racji wewnętrznych” oraz „racji zewnętrznych”, tak jak to zrobiłem w tytule. To, czy istnieją dwa rodzaje racji działania, w odróżnieniu od dwóch rodzajów twierdzeń na temat posiadanych przez ludzi racji działania, jest sprawą do wyjaśnienia. Co więcej, jak się ostatecznie przekonamy, jedna z interpretacji jest problematyczna.

Najpierw rozważę interpretację wewnętrzną i zbadam, jak daleko nas zaprowadzi. Potem, postępując bardziej sceptycznie, zastanowię się, co mogłaby pociągać za sobą interpretacja zewnętrzna. Zakończę kilkoma bardzo krótkimi uwagami łączącymi omawiane tu zagadnienia ze sprawą dóbr publicznych i tzw. problemem gapowiczów.

Najprostszy model reprezentujący interpretację wewnętrzną byłby następujący: *P* posiada rację, żeby  $\alpha$ -ować, wtedy i tylko wtedy, gdy *P* ma jakieś pragnienie, które zostałyby zaspokojone dzięki  $\alpha$ -owaniu. Ewentualnie mogliśmy dokończyć to zdanie w następujący sposób: ... jakieś pragnienie, które zostałyby, jak sądzi *P*, zaspokojone dzięki  $\alpha$ -owaniu; różnicą między tymi wersjami zajmiemy się później. Model taki bywa czasami przypisywany Hume’owi, ale ponieważ poglądy Hume’a są w rzeczywistości bardziej złożone, możemy nazwać go *uproszczonym modelem Hume’owskim*. Jest on z pewnością zbyt prosty. Dokonując w nim poprawek i zmian, zamierzam skonstruować na jego podstawie jakiś bardziej wiarygodny model. Próbując to zrobić, przedstawię cztery sądy, które — jak mi się wydaje — trafnie opisują twierdzenia na temat racji wewnętrznych.

Przede wszystkim każdy model reprezentujący interpretację wewnętrzną na mocy definicji musi pokazać zależność twierdzenia na temat racji od *subiektywnego układu motywacyjnego*, który nazwę *U* podmiotu. Do treści *U* dojdziemy później, na razie możemy powiedzieć tyle:

- (i) Brak odpowiedniego elementu w *U* falsyfikuje twierdzenie na temat racji wewnętrznej.

Najbardziej uproszony model Hume’owski głosi, że jakkolwiek element w *U* jest źródłem racji wewnętrznej. Istnieją jednak podstawy, by temu zaprzeczyć, ale nie ze względu na istnienie w *U* elementów godnych pożałowania, nierozważnych czy dewiacyjnych — one stwarzają innego rodzaju problemy — ale z powodu istnienia w nim elementów opartych na fałszywym przekonaniu.

Podmiot jest przekonany, że dany płyn to gin, gdy w rzeczywistości jest to benzyna. Ma ochotę na gin z tonikiem. Czy posiada rację, jakąś rację, by

zmieszać to coś z tonikiem i wypić? Istnieją tu dwa rozwiązania (na które wskazują alternatywne sformułowania uproszczonego modelu Hume'owskiego). Z jednej strony, bardzo dziwne jest powiedzenie, że posiada on rację, żeby wypić ten płyn; naturalne zaś, że nie posiada on żadnej racji, żeby to zrobić, nawet jeśli uważa, że ją ma. Z drugiej, jeśli wypije ten płyn, mamy nie tylko jakieś wyjaśnienie tego, co zrobił (jakąś rację, dlaczego to zrobił), ale także mamy takie wyjaśnienie, które przyjmuje formę racji działania. Ten wymiar wyjaśniający jest niezwykle istotny i wrócimy do niego jeszcze niejednokrotnie. Jeśli istnieją racje działania, to musi być tak, że ludzie czasami działają ze względu na te racje, a jeśli tak robią, to jakieś trafne wyjaśnienie działań ludzi musi uwzględniać ich racje (co nie znaczy, że wszystkie trafne wyjaśnienia działań ludzi muszą uwzględniać ich racje). Różnica między prawdziwymi a fałszywymi przekonaniem podmiotu nie może zmieniać *formy* wyjaśniania właściwej do opisanego jego działań. Ta okoliczność mogłaby skłonić nas do zignorowania odnotowanej tu intuicji i do uznania, że podmiot mający ochotę na gin, posiada rację, żeby wypić coś, co jest benzyną.

Nie sądzę jednak, że powinniśmy zaakceptować takie rozwiązanie. Poszlibyśmy w złym kierunku, gdybyśmy zgodnie z nim sugerowali, że koncepcja racji wewnętrznej ma związek wyłącznie z wyjaśnianiem, a nie ma nic wspólnego z racjonalnością podmiotu – taka sugestia mogłaby skłaniać do poszukiwania innych rodzajów racji, które byłyby związane z jego racjonalnością. Wbrew temu rozwiązaniu koncepcja racji wewnętrznych jest jednak związana z racjonalnością podmiotu. Jak zobaczymy, to, co możemy mu trafnie przypisać za pomocą twierdzenia w trzeciej osobie na temat racji wewnętrznej, jest tym, co on sam może sobie przypisać w rezultacie namysłu. Tak więc sądzę, że powinniśmy raczej powiedzieć:

- (ii) *PR*, będąc elementem *U*, nie zapewni *P* racji, żeby  $\alpha$ -ować, albo jeśli istnienie *PR* zależy od fałszywego przekonania, albo jeśli przekonanie *P* dotyczące związku między  $\alpha$ -owaniem a zaspokojeniem *PR* jest fałszywe.

(To dwuczęściowe sformułowanie da się zilustrować przykładem ginu i benzyny: *PR* może być rozumiane albo jako pragnienie napicia się tego, co jest w butelce, albo jako pragnienie napicia się ginu). Tak czy inaczej prawdą będzie, że jeśli w tych okolicznościach podmiot  $\alpha$ -uje, to nie tylko istnieje racja, dlaczego  $\alpha$ -ował, ale także ukazuje go to jako działającego – w związku ze swoim fałszywym przekonaniem — racjonalnie.

Możemy odnotować epistemiczną konsekwencję:

- (iii) (a) *P* może mieć fałszywe przekonanie dotyczące twierdzenia na temat racji wewnętrznej odnoszącej się do niego samego, oraz (możemy dodać)
- (b) *P* może nie znać pewnego prawdziwego twierdzenia na temat racji wewnętrznej odnoszącej się do niego samego.

Punkt (b) pochodzi z dwóch różnych źródeł. Pierwszym jest to, że *P* może nie znać pewnego faktu, który jest tego rodzaju, że gdyby go znał, to — ze względu na jakiś element w *U* — byłby skłonny do  $\alpha$ -owania: możemy powiedzieć, że posiada on rację, żeby  $\alpha$ -ować, chociaż o tym nie wie. Jeśli jednak faktycznie ma on posiadać rację, to związek tego nieznanego faktu z działaniami podmiotu musi być dość ścisły i bezpośredni; w innym razie powiedziałoby się jedynie, że *P* posiadałby rację, żeby  $\alpha$ -ować, gdyby znał ów fakt. Nie będę zajmował się warunkami wymaganymi do pierwszego lub drugiego stwierdzenia, ale problem ten musi się ściśle wiązać z pytaniem o to, kiedy niewiedza stanowi część wyjaśnienia tego, co *P* tak naprawdę robi.

Drugim źródłem (iii) jest to, że *P* może nie znać pewnego elementu w *U*. Powinniśmy jednak zauważyć, że ten nieznaną element w *U*, czyli *PR*, zapewni *P* rację, żeby  $\alpha$ -ować, tylko wtedy, gdy  $\alpha$ -owanie jest racjonalnie powiązane z *PR*; to znaczy, z grubsza rzecz biorąc, że plan  $\alpha$ -owania mógł być odpowiedzią na pewien wymagający namysłu problem, który został określony po części przez *PR*. Jeśli *PR* jest nieznaną *P*, ponieważ podmiot jest nieświadomy istnienia w sobie tego elementu, to *PR* równie dobrze może nie spełniać tego warunku, chociaż może oczywiście dostarczać racji, dlatego *P*  $\alpha$ -uje, czyli może wyjaśnić lub pomóc w wyjaśnieniu  $\alpha$ -owania przez *P*. W takich przypadkach  $\alpha$ -owanie może być związane z *PR* tylko symbolicznie.

Właśnie uznałem, że

- (iv) twierdzenia na temat racji wewnętrznych można odkryć podczas namysłu.

Warto wspomnieć o czymś dotąd tylko dorozumianym: twierdzenie na temat racji wewnętrznej ma zastosowanie nie tylko w przypadku tego działania, które zostało wybrane w wyniku namysłu. „*P* posiada rację, żeby  $\alpha$ -ować”, nie znaczy, że „działanie, do którego *P* posiada rację, wzięwszy pod uwagę wszystkie okoliczności, to  $\alpha$ -owanie”. Podmiot może mieć rację, żeby robić wiele rzeczy, w przypadku których będzie miał też inne, silniejsze, racje, by ich nie robić.

Uproszony model Hume’owski zakłada, że  $\alpha$ -owanie musi być związane z pewnym elementem w *U* w taki sposób, w jaki środek wiedzie do celu

(chyba że  $\alpha$ -owanie wprost zaspokaja jakieś pragnienie, które samo jest tym elementem w  $U$ ). To jednak tylko jeden z możliwych przypadków: co więcej, zwykle odkrycie, że jakiś sposób działania jest środkiem wiodącym do celu, nie jest jeszcze, samo w sobie, rozumowaniem praktycznym<sup>2</sup>. Dobrym przykładem rozumowania praktycznego jest rozumowanie zmierzające do konkluzji, że posiada się rację, żeby  $\alpha$ -ować, ponieważ  $\alpha$ -owanie byłoby najbardziej dogodnym, oszczędnym, przyjemnym itd. sposobem na spełnienie pewnego elementu w  $U$ , na co wpływ mają oczywiście też inne elementy w  $U$ , choć niekoniecznie w jasny lub ustalony sposób. Istnieje jednak wiele innych możliwości namysłu: zastanawianie się, w jaki sposób da się z sobą pogodzić spełnianie różnych elementów w  $U$ , np. poprzez ustalanie kolejności ich spełniania; rozważanie, któremu z elementów w  $U$  przypisać największą wagę, w sytuacji, gdy istnieje nierozwiązywalny konflikt między nimi (co istotne, wcale nie znaczy to, że istnieje jakaś jedna własność, którą elementy te posiadają w różnym stopniu); lub wreszcie znajdowanie konstruktywnego rozwiązania, np. decydowanie, w jaki sposób rozerwać się wieczorem – przy założeniu, że to, czego się chce, to właśnie rozrywka.

W wyniku takiego procesu podmiot jest w stanie spostrzec, że posiada rację, żeby coś zrobić, choć wcześniej wcale tego nie dostrzegał. W ten sposób, proces namysłu jest w stanie wywołać nowe działania, dla których istnieją racje wewnętrzne, jak i przywołać nowe racje wewnętrzne dla danych działań. Proces namysłu jest też w stanie sprawić, że pewne elementy znikną z  $U$ . Dzięki refleksji podmiot może uznać, że pewne przekonanie jest fałszywe, a w związku z tym może zrozumieć, że w rzeczywistości nie posiada żadnej racji, żeby coś zrobić, choć wcześniej sądził, że ją ma. Ujmując rzecz bardziej wnikliwie: może sądzić, że posiada rację, żeby zrealizować jakieś przedsięwzięcie, ponieważ nie wysilił wystarczająco swej wyobraźni, by uświadomić sobie konsekwencje tego, co się stanie, gdy faktycznie zrealizuje to przedsięwzięcie. Podczas samodzielnego namysłu, lub namysłu inspirowanego perswazją innych, może wyraźniej uchwycić sens związany z danym przedsięwzięciem i utracić pragnienie zrealizowania go, albo — ujmując rzecz bardziej konstruktywnie — wyobraźnia jest w stanie podsunąć mu nowe możliwości i wykreować nowe pragnienia. (Te możliwości istotne są zarówno w polityce, jak i w działaniu indywidualnym).

<sup>2</sup> Uwaga ta pochodzi od Aurela Kolnaiego – zob. jego „Deliberation is of Ends”, w: *Ethics, Value and Reality* (London, Indianapolis: Hackett Pub., 1978). Zob. też: David WIGGINS, „Deliberation and Practical Reason”, *Proceedings of the Aristotelian Society* 76 (1975-1976): 29–51; częściowo przedrukowane w: Joseph RAZ (red.), *Practical Reasoning* (Oxford: Oxford University Press, 1976).

Nie powinniśmy zatem myśleć o *U* jako o czymś statycznym. Proces namysłu jest w stanie mieć najróżniejszy wpływ na *U* i fakt ten powinien z łatwością dać się pogodzić z teorią racji wewnętrznych. Winna więc ona być bardziej otwarta na różne możliwe elementy w *U*, niż chcieliby tego niektórzy teoretycy. Pierwotnie omawiałem *U*, odwołując się do pragnień, i terminologia ta — podchodząc formalnie — może być użyta do wszystkich elementów w *U*. Terminologia ta jednak może też sprawić, że zapomni się, iż *U* zawiera również takie elementy, jak dyspozycje do oceny, wzory reakcji emocjonalnych, zobowiązania osobiste i różne plany — jak można by to abstrakcyjnie określić — ucieleśniające to, do czego przywiązany jest podmiot. Przede wszystkim nie zakłada się oczywiście, że pragnienia czy plany podmiotu muszą być egoistyczne; można mieć nadzieję, że będzie on miał różnego rodzaju nieegoistyczne plany, które na równi będą mu w stanie dostarczać wewnętrznych racji działania.

Istnieje jednak kolejny problem dotyczący treści *U*: czy powinno się uznać, zgodnie z ogólną koncepcją racji wewnętrznych, że *U* zawiera *potrzeby*. Mówi się, i jest to całkiem naturalne, że *P* posiada rację, żeby dążyć do *X*, wyłącznie na tej podstawie, że potrzebuje on *X*. Ale czy na gruncie teorii racji wewnętrznych wynikanie to jest oczywiste? Istnieje pewien szczególny problem z powyższą tezą, ale tylko pod warunkiem, że możliwe jest niemotywowanie podmiotu przez to, czego on potrzebuje. Nie będę próbował tu rozważać natury potrzeb, przyjmę jednak, że o ile istnieją dające się określić potrzeby, o tyle może też istnieć podmiot, który nie przejawia żadnego zainteresowania uzyskaniem tego, czego naprawdę potrzebuje. Przyjmę ponadto, iż namysł nie wpływa na ten brak zainteresowania, a także, że nie wolno twierdzić, iż to niezainteresowanie zawsze musi być wynikiem fałszywego przekonania. (A jeśli rzeczywiście jest wynikiem fałszywego przekonania, to podpada pod przypadek (ii) w omówiony powyżej sposób).

Jeśli podmiot rzeczywiście nie jest zainteresowany dążeniem do tego, czego potrzebuje — i nie jest to wynik fałszywego przekonania, a także podmiot nie byłby w stanie osiągnąć takiego motywu, wychodząc z posiadanych przez siebie motywów za pomocą omówionych przez nas metod namysłu — to uważam, że musimy powiedzieć, iż w sensie wewnętrznym podmiot faktycznie nie posiada żadnej racji, żeby dążyć do tego, czego potrzebuje. Niemniej, mówiąc to, musimy pamiętać, jak mocne są te założenia i jak rzadko bylibyśmy skłonni uważać, że wiemy, iż są prawdziwe. Kiedy mówimy, iż jakaś osoba posiada rację, żeby zażyć potrzebny jej lek, choć ciągle i przekonująco zaprzecza, że jest zainteresowana zachowaniem zdrowia, to

wciąż możemy to mówić w sensie wewnętrznym, mając na myśli, że tak naprawdę na pewnym poziomie ona *musi* chcieć czuć się dobrze.

Jeśli jednak wyraźnie byśmy uznawali, że niczego takiego nie mamy na myśli, a jednocześnie nalegali, iż ta osoba posiada rację, żeby zażyć lekarstwo, to wówczas musielibyśmy wypowiadać się w innym sensie — w sensie zewnętrznym. Ludzie mówią rzeczy, które faktycznie aż się proszą o interpretację zewnętrzną. W noweli Henry’ego Jamesa na temat Owena Wingrave’a, na której podstawie Britten stworzył operę, rodzina Owena namawia go do wstąpienia do armii, uznając, że jest to konieczne i ważne, ponieważ wszyscy jego mężczyźni przodkowie byli żołnierzami, a honor rodzinny wymaga od niego tego samego. Owen Wingrave nie ma żadnej motywacji, by wstąpić do armii, a wszelkie jego pragnienia kierują go w przeciwną stronę: nienawidzi wszystkiego, co wiąże się z żołnierskim życiem. Jego rodzina mogła była wyrazić swoje życzenie, mówiąc, że *istnieje dla Owena racja, żeby wstąpić do armii*. Wiedząc, że w *U* Owena nie było nic, co — poprzez namysł — skłoniłoby go do realizacji tych planów, nie wycofali się ze swojego twierdzenia ani nie przyznaliby, że tkwili w błędnym mniemaniu. Mieli oni na myśli zewnętrzny sens. Jaki jest to sens?

Przede wszystkim trzeba stwierdzić, że nie jest to ten sam problem, który dotyczy statusu domniemanego imperatywu kategorycznego w kantowskim sensie „powinności”, mającym zastosowanie niezależnie od tego, czego podmiot akurat chce — a raczej nie jest to bez żadnych wątpliwości ten sam problem. Po pierwsze, imperatyw kategoryczny często był interpretowany, choćby przez Kanta, jako imperatyw z konieczności związany z moralnością, tymczasem twierdzenia na temat racji zewnętrznych niekoniecznie odnoszą się do moralności. Po drugie, nie jest jasne, jaki jest związek między „istnieje dla *P* racja, żeby...” a „*P* powinien...”. Niektórzy filozofowie uważali te wyrażenia za równoważne i przy takim rozumieniu problem racji zewnętrznych jest oczywiście znacznie bliższy zagadnieniu imperatywu kategorycznego. Niemniej nie będę czynił żadnych założeń na temat tej równoważności ani omawiał kwestii „powinności”<sup>3</sup>.

Rozważając, co twierdzenie na temat racji zewnętrznej mogłoby znaczyć, znów musimy pamiętać o formie możliwego wyjaśnienia, czyli o jakimś względnie odnoszącym się do każdej racji działania. Jeśli coś może być racją działania, to znaczy, że przy konkretnej okazji mogłoby być czyjąś racją

<sup>3</sup> Jest ona dyskutowana w rozdziale 9 [książki Bernarda Williama *Moral Luck: Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981), w której przedrukowano niniejszy tekst — przyp. Tłum.].

działania i byłoby wtedy uwzględnione w wyjaśnieniu tego działania. Tymczasem żadne twierdzenie na temat racji zewnętrznej nie byłoby w stanie *samo w sobie* dostarczyć wyjaśnienia czyjegoś działania. Nawet jeśli byłoby prawdą (cokolwiek może to znaczyć), że istniała dla Owena racja, żeby wstąpić do armii, to fakt ten sam w sobie w ogóle nie wyjaśniłby niczego, co Owen zrobił, nawet jego wstąpienia do armii. Jeśli to twierdzenie na temat racji zewnętrznej byłoby w ogóle prawdziwe, to byłoby prawdziwe także wtedy, gdyby Owen nie był motywowany do wstąpienia do armii. Sedno problemu z twierdzeniami na temat racji zewnętrznych polega na tym, że mogą one być prawdziwe niezależnie od motywacji podmiotu. Tymczasem nic nie jest w stanie wyjaśnić (zamierzonych) działań podmiotu za wyjątkiem czegoś, co motywuje go do działania. Tak więc, by wyjaśnić działanie, trzeba czegoś więcej poza prawdziwością twierdzenia na temat racji zewnętrznej, trzeba jakiegoś psychologicznego związku, którym wydaje się być przekonanie. Przekonanie *P* dotyczące twierdzenia na temat zewnętrznej racji odnoszącej się niego samego może pomóc w wyjaśnieniu jego działania.

Twierdzenia na temat racji zewnętrznych zostały wprowadzone jedynie w ogólnej formie: „istnieje racja dla *P*, żeby...”, ale teraz potrzebujemy wyjść poza tę formułę i przejść do konkretnych twierdzeń na temat racji. Bez wątpienia istnieją przypadki  $\alpha$ -owania przez podmiot, ponieważ jest on przekonany, że istnieje dla niego racja, żeby  $\alpha$ -ować, choć nie ma on żadnych przekonań na temat tego, jaka ta racja jest. Byłyby to przypadki polegania na autorytecie, któremu podmiot ufa, albo przypadki, w których przypomina on sobie, że znał jakąś rację za  $\alpha$ -owaniem, ale teraz nie jest w stanie jej sobie przywołać. Pod tym względem racje działania są jak racje za przekonaniem. Podobnie jednak jak w przypadku tych drugich racji są to wyraźnie wtórne przypadki. Podstawowy przypadek musi być taki, że *P*  $\alpha$ -uje nie dlatego, że jest przekonany wyłącznie o istnieniu dla niego takiej czy innej racji, żeby  $\alpha$ -ować, lecz dlatego, że jest przekonany o istnieniu pewnych określonych względów, które konstytuują dla niego rację, żeby  $\alpha$ -ować. Tak więc mogłoby dojść do tego, że Owen Wingrave wstąpiłby do armii, ponieważ byłby (w takim wypadku) przekonany o istnieniu dla siebie racji, żeby to zrobić — takiej mianowicie, że jego rodzina ma tradycję służby wojskowej.

Czy przekonanie, że pewien określony wzgląd jest racją, żeby działać w pewien określony sposób, zapewnia lub, co więcej, konstytuuje motywację do działania? Jeśli nie, to nie zrobiliśmy ani kroku w przód. Załóżmy więc, że tak — to twierdzenie faktycznie wydaje się przekonujące, przynajmniej o tyle, o ile związek między takimi przekonaniem a dyspozycją działania nie zostanie wzmocniony tak nadmiernie, by wykluczyć *akrazję*.



Twierdzenie to jest w rzeczywistości *tak* wiarygodne, że ów podmiot posiadający to przekonanie wydaje się być teraz jednym z tych, o których można by prawdziwie orzec twierdzenie na temat racji *wewnętrznej*: jest kimś, kto ma w swoim *U* właściwą motywację. Człowiek, który naprawdę jest przekonany, że względy dotyczące honoru rodziny konstytuują rację działania, jest człowiekiem z pewną dyspozycją do działania, a także z dyspozycjami do aprobaty, uczucia, reakcji emocjonalnej i tak dalej.

Z tego nie wynika, że twierdzenia na temat racji zewnętrznych są puste. Wynika natomiast, że ich treści nie odkryje się ani poprzez rozważanie wyłącznie stanu tego kogoś, kto uznaje takie twierdzenie, ani poprzez to, w jaki sposób jego stan wyjaśnia działanie, ponieważ o stanie tym można prawdziwie orzec jedynie twierdzenie na temat racji wewnętrznej. Zamiast tego treść twierdzeń typu zewnętrznego musi zostać odkryta poprzez rozważenie, jak to jest *uznać* twierdzenie — a to znaczy, że ich cecha charakterystyczna będzie się musiała dopiero wyłonić.

Zajmiemy się przykładem (co implicytnie już robimy), w którym twierdzenie na temat racji zewnętrznej odnosi się do kogoś takiego jak Owen Wingrave, kto nie jest jak dotąd motywowany w wymagany sposób, a więc jest kimś, w stosunku do kogo nie można by prawdziwie użyć twierdzenia na temat racji wewnętrznej. (Ponieważ różnica między twierdzeniami na temat racji wewnętrznych i zewnętrznych zależy od założeń mówiącego, to twierdzenia na temat racji zewnętrznych da się oczywiście odnieść do podmiotów, które są już motywowane; nie jest to jednak interesujący przypadek). Podmiot nie uznaje obecnie twierdzenia na temat racji zewnętrznej. Jeśli uzna takie twierdzenie, będzie motywowany do działania; tak więc uznanie twierdzenia musi w zasadzie wiązać się z nabyciem nowej motywacji. Ale jak to jest możliwe?

Problem ten jest blisko związany ze starym pytaniem, jak „rozum może być źródłem motywacji”, pytaniem, na które Hume udzielił swej słynnej negatywnej odpowiedzi. W tej jednak formie pytanie to samo w sobie jest niejasne i w niejasny sposób wiąże się z naszą argumentacją — ponieważ rozum, to znaczy racjonalne procesy, może oczywiście być źródłem nowych motywacji, jak widzieliśmy to przy okazji wyjaśniania namysłu. Co więcej, tradycyjny sposób przedstawiania problemu (na co będę wskazywał) bierze na siebie obowiązek wskazania, co ma się liczyć jako „czysto racjonalny proces”, gdy tymczasem nie powinien podejmować się tego zadania, bo obowiązek ten faktycznie spoczywa na tym, kto chce zaprzeczyć ogólnej konkluzji Hume’a i kto przypisuje twierdzeniom na temat racji zewnętrznych ważną rolę, czyli na kimś, kogo będę nazywał „teoretykiem racji zewnętrznych”.

Sednem sprawy jest dostrzeżenie, że teoretyk racji zewnętrznych musi w *szczególny sposób* pojmować związek między nabyciem motywacji a uznaniem twierdzenia na temat racji. Istnieją oczywiście różne środki, za których pomocą podmiot mógłby nabyć motywację, a także uznać twierdzenie na temat racji, ale dla teoretyka racji zewnętrznych nie są one odpowiednie. Poruszająca retoryka rodziny Owena mogłaby go do tego stopnia przekonać, że zarówno nabyłby motywację, jak i uznałby twierdzenie. Wyklucza to jednak element, na którym teoretykowi racji zewnętrznych zależy przede wszystkim, czyli to, że podmiot powinien nabyć nową motywację, *ponieważ* uznaje twierdzenie na temat racji zewnętrznej, a także to, że powinien je uznać, ponieważ pod pewnym względem pojmuje sprawy właściwie. Jeśli teoretyk ten zamierza trzymać się tych wymogów, to będzie musiał — jak sądzę — ustalić warunek, pod którym podmiot we właściwy sposób nabywa motywację tego typu, że powinien rozumować poprawnie; a samo twierdzenie na temat racji zewnętrznych będzie musiało być traktowane jako z grubsza równoznaczne lub — co najmniej — jako pociągające za sobą twierdzenie, że jeśli podmiot dokonałby racjonalnego namysłu, to wtedy — bez względu na to, jakie motywacje posiadałby pierwotnie — nabyłby motywację, żeby  $\alpha$ -ować.

Jeśli jednak o rozwiązanie jest trafne, to istotnie wydaje się, że w głównej tezie Hume'a jest sporo racji, a także że jest bardzo prawdopodobne, iż wszystkie twierdzenia na temat racji zewnętrznych są fałszywe. A to z tego względu, że *ex hypothesi* nie istnieje w podmiocie żadna motywacja, *na której podstawie* dokonywałby on namysłu, by nabyć tę nową motywację. Biorąc pod uwagę zarówno wcześniej istniejące motywacje podmiotu, jak i tę nową, do uznania, że twierdzenia na temat racji zewnętrznych są prawdziwe — na mocy tej interpretacji — potrzebne jest, by do tej nowej motywacji można by dojść w pewien racjonalny sposób, zakładając wcześniejsze motywacje. Jednocześnie nie wolno jej mieć z wcześniejszą motywacją żadnego takiego racjonalnego związku, który rozważaliśmy powyżej, omawiając namysł — ponieważ w takim wypadku to przede wszystkim twierdzenie na temat racji wewnętrznej byłoby prawdziwe. Nie widzę żadnej racji, by zakładać, że te warunki mogłyby zostać spełnione.

Można by rzec, że siłę twierdzenia na temat racji zewnętrznej da się wyjaśnić w następujący sposób. Twierdzenie tego typu implikuje, że racjonalny podmiot byłby motywowany, by działać właściwie, a jest ono w stanie mieć taką implikację, ponieważ racjonalny podmiot to właśnie ktoś, kto ma w swoim  $U$  ogólną dyspozycję do robienia tego, do czego (jak sądzi) istnieje dla niego racja. Tak więc kiedy uznałby, że istnieje dla niego racja, żeby  $\alpha$ -ować,

to byłby motywowany do  $\alpha$ -owania, nawet jeśli wcześniej nie miałby ani motywacji, żeby  $\alpha$ -ować, ani żadnej innej motywacji związanej — na jeden ze sposobów, które rozważaliśmy, omawiając namysł — z  $\alpha$ -owaniem.

Ta jednak odpowiedź odsuwa jedynie problem. Z grubsza rzecz biorąc, ponownie wykorzystuje model pragnień i przekonań do wyjaśnienia działania, lecz używa jakiegoś pragnienia i jakiegoś przekonania, których treść należy dopiero wyjaśnić. Co takiego sprawia, że ktoś dochodzi do przekonania w sytuacji, w której uznaje, że istnieje dla niego racja, żeby  $\alpha$ -ować, jeśli nie jest to taki sąd — lub coś, co pociąga za sobą taki sąd — że gdyby dokonał racjonalnego namysłu, to byłby motywowany, by działać właściwie? Pytaliśmy, w jaki sposób jakikolwiek prawdziwy sąd mógłby mieć taką treść; w uzyskaniu odpowiedzi na to pytanie nie jest nam w stanie pomóc odwołanie się do rzekomego pragnienia wywoływanego przez przekonanie mające taką właśnie treść.

Powyższe argumenty, odnoszące się do tego, jak to jest uznać twierdzenie na temat racji zewnętrznej, wiążą się pewną koncepcją dotyczącą tego, co przedstawione tu wyjaśnienie procesu namysłu dopuszcza, a co wyklucza. W tym miejscu ktoś może jednak zarzucić nam, że wyjaśnienie procesu namysłu jest bardzo mętne i że na przykład dopuszcza wykorzystywanie wyobraźni w celu rozszerzenia lub ograniczenia treści  $U$  podmiotu. A jeśli faktycznie tak jest, to nie jest jasne, jakie są granice tego, do czego poprzez racjonalny namysł może dojść podmiot, wychodząc ze swego obecnego  $U$ .

To *jest* niejasne i uważam, że to niezwykle pożądana cecha teorii rozumowania praktycznego, cecha, którą teoria ta powinna zachować i objaśnić. Istnieje zasadnicza nieokreśloność tego, co da się uznać za proces racjonalnego namysłu. Rozumowanie praktyczne jest heurystycznym procesem wymagającym wyobraźni i nie istnieją żadne określone granice na kontinuum znajdującym się między racjonalną myślą a inspiracją lub nawróceniem. Dla kogoś, kto uważa, że racje działania mają być zasadniczo rozumiane wedle modelu wewnętrznego, nie jest to problem. W istocie niejasność dotycząca „ $P$  posiada rację, żeby  $\alpha$ -ować” w wewnętrznym sensie wynika z tego, że procesy namysłu, które mogłyby prowadzić  $P$  od obecnego  $U$  do bycia motywowanym, żeby  $\alpha$ -ować, mogą być mniej lub bardziej szeroko pojęte. Nie jest to jednak żaden kłopot dla tych, którzy za podstawową uznają wewnętrzną koncepcję racji działania. Sytuacja ta pokazuje jedynie, że istnieje szerszy i w mniejszym stopniu określony, niż można by było przypuszczać, przedział stanów, które mogą być traktowane jako racje dla  $P$ , żeby  $\alpha$ -ować.

To teoretyk racji zewnętrznych ma w tym momencie problem. Istnieje oczywiście wiele rzeczy, które mówiący może powiedzieć komuś, kto nie

jest skłonny do  $\alpha$ -owania, a kto, zdaniem mówiącego, powinien być do tego skłonny — choćby to, że jest nierozważny, okrutny, samolubny, nierozsądny; lub też, że sprawy, a także on sam, miałyby się znacznie lepiej, gdyby miał taką motywację. Powiedzenie którejkolwiek z tych rzeczy może być sensowne. Ktoś jednak, kto wkłada wiele wysiłku w krytykę z pozycji racji zewnętrznych, wydaje się twierdzić, że to, co jest szczególnie złe w tym podmiocie, to to, że jest on *irracjonalny*. Dlatego to właśnie teoretyk racji zewnętrznych jest tym, który musi sprecyzować swoje oskarżenie: zwłaszcza, że to on od każdego racjonalnego podmiotu, jako takiego, żąda uznania wymogu zrobienia danej rzeczy.

Rodzina Owena Wingrave'a nie posługiwała się wyrażeniem „racje”, lecz — jak sobie wyobraziliśmy — mogłaby użyć formuły odwołującej się do racji zewnętrznych. Sam ten fakt sprawia teoretykowi racji zewnętrznych pewne trudności. Teoretyk ten, który traktuje prawdziwość twierdzenia na temat racji zewnętrznej jako potencjalną podstawę oskarżenia tego, kto ignoruje ową rację, o irracjonalność, może równie dobrze powiedzieć, że jeśli Wingrave'owie wyłożyliby swe zastrzeżenia względem Owena w tej formie, to bardzo prawdopodobne byłoby, że twierdziliby coś, co — w tym konkretnym przypadku — było fałszywe. Kwestia, która sprawiłaby teoretykowi racji zewnętrznych jeszcze więcej trudności, polegałaby na pokazaniu, że słowa *znaczą* coś innego, kiedy używane są przez Wingrave'ów, od tego, co znaczą, kiedy wypowiedane są — jak przypuszcza teoretyk — trafnie. To jednak, co znaczą, kiedy wypowiedają je Wingrave'owie, to prawie na pewno *nie* to, że racjonalny namysł doprowadziłby Owena do nabycia motywacji, żeby wstąpić do armii — a takie (z grubsza rzecz biorąc) znaczenie lub implikację przypisaliśmy tym słowom, jeśli mają być one tak ważne, jak chcieliby tego zwolennicy tej teorii.

Przedstawione tu względy zdecydowanie przemawiają, według mnie, za tym, że twierdzenia na temat racji zewnętrznej, gdy traktować je jako ściśle wyodrębnione, są fałszywe, niespójne albo też są czymś, co ma wprowadzać w błąd. W rzeczywistości jednak znacznie ciężiej jest wyodrębnić je w wypowiedziach ludzi, niż sugerowało to ich wprowadzenie na początku tego artykułu. Ci, którzy posługują się tego rodzaju określeniami, często zdają się formułować raczej konkretną tezę na temat racji wewnętrznej, choć faktycznie niekiedy twierdzenie takie traktuje się jakby pochodziło zdecydowanie spoza *U* podmiotu, a także spoza tego, co mógł on wywieść z *U* za pomocą racjonalnego namysłu — a wówczas, jak sądzę, jest bardzo niejasne, co takie twierdzenie znaczy. Czasami znaczy ono jedynie niewiele więcej

niż to, że sprawy miałyby się lepiej, gdyby podmiot postąpił tak, a nie inaczej. Jednakże sformułowanie odwołujące się do racji faktycznie przynosi pewien skutek, szczególnie gdy sugeruje, że podmiot jest irracjonalny — ale sugestia ta, jeśli odrzuca się jako jej podstawę jakieś twierdzenie na temat racji wewnętrznej, jest blefem. Jeśli tak jest, to jedynymi prawdziwymi twierdzeniami na temat racji działania będą twierdzenia wewnętrzne.

Uważa się, że z dyskutowanym tu zagadnieniem bardzo blisko związane są problemy dóbr publicznych i gapowiczów dotyczące sytuacji — z grubsza rzecz biorąc — gdy każda osoba ma egoistyczną rację, by chcieć pewnego dostarczanego dobra, lecz jednocześnie każdy posiada egoistyczną rację, by nie brać udziału w dostarczaniu tego dobra. Nie będę próbował omawiać tego problemu, lecz pomocne może być przedstawienie na koniec — po prostu, by pokazać mój własny pogląd na temat racji działania oraz uwidocznić różnicę z pewnymi innymi stanowiskami — listy pytań mających związek z tym zagadnieniem, wraz z odpowiedziami, których udzieliłby ten, kto uważa, że jedyna racjonalność działania to racjonalność racji wewnętrznych.

1. Czy jesteśmy w stanie zdefiniować koncepcje racjonalności, które nie są czysto egoistyczne?

Tak.

2. Czy jesteśmy w stanie zdefiniować koncepcje racjonalności, które nie są czysto instrumentalne?

Tak.

3. Czy jesteśmy w stanie zdefiniować koncepcję racjonalności, wedle której działanie racjonalne dla  $P$  nie jest w żaden sposób związane z istniejącymi motywacjami  $P$ ?

Nie.

4. Czy jesteśmy w stanie wykazać, że osoba, która ma jedynie egoistyczne motywacje jest irracjonalna, nie dążąc do nieegoistycznych celów?

Niekoniecznie, chociaż możemy być w stanie to zrobić w szczególnych wypadkach. (Problem z osobą egoistyczną zazwyczaj nie polega na tym, że jest ona akurat irracjonalna).

Założmy, że istnieje jakieś dobro  $D$  i taki zbiór osób  $O$ , że każdy członek  $O$  posiada egoistyczną rację, żeby chcieć  $D$ , ale dostarczenie  $D$  wymaga działania  $DZ$ , z którym wiążą się koszty dla każdej jednostki z pewnego podzbioru  $O$ ; założmy też, że  $P$  jest członkiem  $O$ , a wtedy:

1. Czy  $P$  posiada egoistyczną rację, żeby wykonać  $DZ$ , jeśli w rozsądny sposób zakłada, albo że zbyt mało członków  $O$  wykona  $DZ$ , by  $D$  zostało

dostarczone, albo że wystarczająco wielu pozostałych członków *O* wykona *DZ*, tak że dobro *D* zostanie dostarczone?

Nie.

2. Czy istnieją jakieś takie warunki, w których *P* jest w stanie mieć egoistyczną rację, żeby wykonać *DZ*?

Tak, w tych przypadkach, w których osiągnięcie krytycznej liczby wykonujących *DZ* zależy od wykonania *DZ* przez *P*, lub od tego, czy posiada on rację, by uważać, że tak jest.

3. Czy istnieją jakieś motywacje, które sprawiłyby, że byłoby racjonalnie dla *P* zrobić *DZ*, choć akurat nie w opisaną wyżej sytuacji?

Tak, a jeśli *P* nie jest całkowitym egoistą, to wiele. Na przykład istnieją motywacje ekspresywne – występujące np. w głośnym przypadku głosowania<sup>4</sup>. Istnieją również motywacje, które wywodzą się z poczucia bezstronności. Wtedy da się wyjść poza dylemat: „albo bezużyteczne, albo niekonieczne”, stosując argument tego typu: „ktoś, ale skoro nie istnieje racja, by pomijać kogokolwiek, to każdy”.

4. Czy jest irracjonalne dla podmiotu, by miał takie motywacje?

W jakimkolwiek sensie, w jakim to pytanie jest zrozumiałe, nie.

5. Czy jest racjonalne dla społeczeństwa, by wychowywało ludzi mających takie motywacje?

Na ile to pytanie jest zrozumiałe, tak. A z pewnością posiadamy racje, by zachęcać ludzi do posiadania tych dyspozycji — np. z racji posiadania przez nas samych tych motywacji.

Przyznaję, że nie dostrzegam żadnych innych poważnych pytań, które na tym poziomie ogólności mają znaczenie dla omawianych tu kwestii. Wszystkie powyższe pytania mają jasne odpowiedzi, które są w pełni zgodne z koncepcją racjonalności praktycznej opartą na wewnętrznych racjach działania, i które, jak mi się wydaje, są całkiem rozsądnie uzasadnione.

*Z języka angielskiego przełożył Tomasz Żuradzki*

<sup>4</sup> Dobrze znane opracowanie przedstawił Mancur Olson, Jr. w *The Logic of Collective Action* (Cambridge, Mass.: Harvard University Press, 1965). Na temat ekspresywnych motywacji w tym znaczeniu zob. Stanley I. BENN, „Rationality and Political Behaviour”, w: Stanley I. BENN i G.W. MORTIMORE (red.), *Rationality and the Social Sciences* (London: Routledge, 1976). Oczywiście można znacznie więcej powiedzieć na temat wynikającego z tekstu problemu bezstronności — na przykład, w jaki sposób członkowie grupy są w stanie, zgodnie z wymogiem bezstronności, zgodzić się na strategię efektywniejsze niż wykonywanie *DZ* przez każdego (tj. przykładowo na wykonywanie *DZ* po kolei).

Dr TOMASZ ŻURADZKI — pracownik Instytutu Filozofii UJ, kierownik Interdyscyplinarnego Centrum Etyki UJ; adres do korespondencji: Instytut Filozofii UJ, ul. Grodzka 52, 31-044 Kraków, e-mail: t.zuradzki@uj.edu.pl. ORCID: <https://orcid.org/0000-0001-6749-2124>.

## RACJE WEWNĘTRZNE I ZEWNĘTRZNE

## Streszczenie

Artykuł, opublikowany po raz pierwszy w 1979 r., jest jednym z najczęściej cytowanych tekstów filozoficznych z drugiej połowy XX wieku. Tekst Bernarda Williamsa zainicjował kilka ważnych debat, toczących się do dziś w etyce i filozofii działania. Zaproponowana przez niego interpretacja pojęcia racji działania jest, z jednej strony, niezwykle wpływowa, ale z drugiej bardzo niejednoznaczna i często krytykowana. Williams broni stanowiska, które z czasem zaczęto określać jako internalizm racji: pewne względy są racjami działania dla danego podmiotu tylko wtedy, gdy mają ścisły związek z „subiektywnym układem motywacyjnym” tego podmiotu, czyli z jakimiś aspektami jego psychiki, charakteru, celów, pragnień, planów, relacji z innymi itd. Stwierdzenie, że ktoś ma rację, by coś zrobić (a jeśli uznajemy, że jest to racja przeważająca — to, że powinien to coś zrobić), znaczy więc wedle tego stanowiska tyle, że ten ktoś jest w stanie być motywowany, by to coś zrobić. W przeciwieństwie do tego stanowisko określane obecnie jako eksternalizm racji głosi, że pewne względy mogą być racjami działania dla danego podmiotu także wtedy, gdy odwołują się do niezależnych od jego układu motywacyjnego własności świata. Williams wychodzi od modelu działania nawiązującego do koncepcji Hume’owskiej, następnie przedstawia jego bardziej rozbudowaną wersję, omawia rolę namysłu, naturę potrzeb, a także przedstawia argument za internalizmem racji, określany jako argument z wyjaśniania działania. Głosi on, że nawet gdyby istniały zewnętrzne racje działania, niezwiązane z subiektywnym układem motywacyjnym danego podmiotu, to odwołanie się do nich nie byłoby w stanie wyjaśnić, dlaczego dany podmiot postąpił tak, a nie inaczej, gdyż „nic nie jest w stanie wyjaśnić (zamierzonych) działań podmiotu za wyjątkiem czegoś, co motywuje go do działania”. Rozważania na temat tego argumentu są dla Williamsa okazją m.in. do przedstawienia własnej interpretacji głośnej tezy Davida Hume’a, głoszącej, że rozum nie może być źródłem motywacji do działania.

## INTERNAL AND EXTERNAL REASONS

## Summary

The paper, originally published in 1979, is one of the most cited philosophical texts written in the second half of the 20<sup>th</sup> century and initiated a series of important discussions in ethics and the philosophy of action that continue to be of relevance in the present. Bernard Williams’ interpretation of the concept of internal reasons is, on the one hand, extraordinarily influential but very ambiguous and highly criticized on the other. Williams defends the view that was later termed reasons internalism: some considerations are reasons for actions for an agent only if they are closely linked with “the subjective motivational set” of an agent, i.e. with some aspects of his psyche, character, aims, desires, projects, relations with other etc. According to this view, saying that an agent has a reason to do something (or if one thinks that this is a prevailing reason — that an agent ought to do something) means that the agent could be motivated to do it. In contrast, reasons externalism claims that some considerations may be reasons for action for an agent, even if they refer to some things independent of his motivation set. Williams starts with the simplified Humean model of motivation, then builds on it a more adequate model, he discusses the role of deliberation, the nature of needs, and then he presents an argument for reasons internalism. The argument from explanation claims that even if there existed external reasons for action unconnected with a subjective motivation set of an agent, appealing to them could not explain why an agent acts in such a way, because “nothing can explain an agent’s (intentional) actions except

something that motivates him so to act.” The discussion around this argument gives Williams the opportunity to present his own interpretation of the well-known thesis by David Hume that reason cannot give rise to a motivation for action.

**Słowa kluczowe:** Bernard Williams; normatywność; racje działania; internalizm; David Hume.

**Key words:** Bernard Williams; normativity; reasons for action; internalism; David Hume.

**Information about Author:** Prof. BERNARD WILLIAMS (1929–2003) — British ethicist and philosopher; he was a lecturer at the universities of Cambridge, Oxford, and California at Berkeley. Author among others: *Problems of the Self* (1973), *Utilitarianism: For and Against* (1973 — with J.J.C. Smart), *Ethics and The Limits of Philosophy* (1985), *Shame and Necessity* (1993). His books appeared in Polish: *Ile wolności powinna mieć wola* (1999), and *Moralność. Wprowadzenie do etyki* (2001).

**Information about Translator:** TOMASZ ŻURADZKI, PhD — assistant professor at the Institute of Philosophy at the Jagiellonian University; director of the Interdisciplinary Center for Ethics at the Jagiellonian University; address for correspondence: Instytut Filozofii UJ, ul. Grodzka 52, 31–044 Kraków; e-mail: t.zuradzki@uj.edu.pl. ORCID: <https://orcid.org/0000-0001-6749-2124>.