

Polemiki

DOBRO I ZŁO W ŚWIETLE LOGIKI PREFERENCJI *

Zadaniem niniejszych uwag jest wysunięcie pewnych zarzutów pod adresem definicji pojęcia dobra (i zła) sformułowanej przez Georga Henrika von Wrighta w jego znanej książce *The Logic of Preference* (s. 68), a także sformułowanie — w terminach Wrightowskiej logicznej teorii preferencji — bardziej, jak mi się zdaje, zadowalających definicji tych pojęć. Z góry trzeba zastrzec, że język logiki preferencji von Wrighta jest stosunkowo ubogi; nie zakłada się w nim np. żadnych funktorów asercyjnych (w rodzaju: X uznaje zdanie „ p ” za prawdziwe, czy też: X jest przekonany, że p). Z tego względu każda definicja pojęcia dobra (czy zła) sformułowana w języku tej teorii będzie nieadekwatna: ludzie wszak preferują nie rzeczywiste stany rzeczy, lecz stany rzeczy, o których sądzą, iż są rzeczywiste, czyli stany rzeczy, o których zachodzeniu są przekonani. Usunięcie jednak założenia, iż podmiot preferujący ma doskonałą wiedzę, wiedzę zawierającą wyłącznie zdania prawdziwe, które jest milcząco przyjęte w książce von Wrighta, doprowadziłoby do znacznego skomplikowania języka tej teorii. Należałoby język logiki preferencji wzbogacić o któryś z funktorów asercyjnych, przyjmując charakteryzujące go aksjomaty i potem dopiero wprowadzić swoiste postulaty dla pojęcia preferencji. Z tego względu poprzestanę na „wewnętrznej” krytyce definicji podanych przez von Wrighta i „wewnętrznym”, by tak rzec, ich rozwinięciu, tj. na próbie ich modyfikacji w języku używanym przez tego autora.

I

Pojęcie preferencji P jest pierwotnym pojęciem teorii von Wrighta. Występuje ono w kontekstach typu pPq , gdzie p, q to stany rzeczy, i jest stale zrelatywizowane do określonego podmiotu.

Intuicje związane z pojęciem preferowania bezwarunkowego objaśnia autor przyjmując, iż wyrażenie pPq znaczy, że podmiot woli taki świat, w którym jest tak, że $(p \cdot \sim q)$ (tzn. p i nie- q) niż taki świat, w którym jest tak, że $(\sim p \cdot q)$; za jego pomocą definiuje pojęcie dobra: stan rzeczy p jest dobry (dla podmiotu

* W związku z książką G. H. von Wrighta, *The Logic of Preference*, Edinburgh 1963.

w danym czasie), gdy jest on bezwarunkowo preferowany (przez dany podmiot w danym czasie) przed stanem rzeczy $\sim p$. Definicja ta nie wydaje się wszakże adekwatna merytorycznie. Wszak skazaniec, który ma do wyboru śmierć przez powieszenie i śmierć w wyniku długotrwałych tortur i który preferuje bezwarunkowo śmierć przez powieszenie, nie uznałby zapewne tego rodzaju śmierci za dobro. Wypadek ten można uogólnić przez wprowadzenie pojęcia sytuacji przymusowej. Otóż powiemy, że w sytuacji przymusowej znajduje się ktoś, kto wie, że będzie tak, że $(p + q + \dots + z)$, przy tym preferuje on bezwarunkowo $\sim p$ przed p , $\sim q$ przed q , ..., $\sim z$ przed z , a ponadto jego relacja preferencji P jest spójna w zbiorze $\{p, q, \dots, z\}$. Nawet jeśli jest tak, że ów podmiot preferuje któryś ze stanów rzeczy p, q, \dots, z przed wszystkimi pozostałymi stanami rzeczy (czyli preferuje np. p przed $(q + \dots + z)$), to i tak p nie staje się dlań dobrem w intuicyjnym sensie tego słowa. Najprostsza modyfikacja definicji podanej przez autora polegałaby przeto na przyjęciu, iż p jest dobrem (dla osoby X w czasie t), gdy jest preferowane bezwarunkowo przed $\sim p$ (przez X -a, w t), i X nie znajduje się w tym czasie w sytuacji przymusowej. Ale i ta definicja nie zabezpiecza przed zarzutem nieadekwatności, jako że każde uznać za dobro dla sekretarza sądowego to, iż sprawy wpływające w dni parzyste dają do rozpatrzenia sędziemu A (a sprawy wpływające w dni nieparzyste — sędziemu B), choć równie dobrze mógłby podjąć decyzję przeciwną, podjął zaś tę, bo jakąś podjąć musiał. Zarzut ten godzi w koncepcję autora (zarówno oryginalną, jak i wyżej zmodyfikowaną), bo choć przyjmuje on, iż jeśli pPq , to nie wynika stąd, iż dokonany został przez podmiot wybór między p a q , to jednak leży w jego intencjach, by zachodziła zależność odwrotna: stąd, że dokonany został wybór między p a q wynika, że pPq (s. 15 i n.).

II

Z tych względów wydaje się, iż definicja pojęcia dobra w języku logiki preferencji winna być sformułowana inaczej. Niech będzie dany zbiór stanów rzeczy U przebiegany przez zmienne „ p ”, „ q ”, „ r ”, ... Autor przyjmuje, że relacja P jest asymetryczna:

$$(a) \quad pPq \rightarrow \sim (qPp)$$

i przechodnia:

$$(p) \quad (pPq) \cdot (qPr) \rightarrow (pPr)$$

(gdzie „ \rightarrow ” czytamy: „jeżeli, ... to”). Znaczy to więc, że relacja P częściowo porządkuje uniwersum dyskursu U . Możemy przeto zdefiniować relację obojętności N :

$$pNq, \text{ gdy } \sim (pPq) \cdot \sim (qPp),$$

która jest też przechodnia:

$$(pNq) \cdot (qNr) \rightarrow (pNr).$$

Widać, że relacja N jest relacją równościową, tzn. zwrotną:

$$pNp,$$

symetryczną:

$$pNq \rightarrow qNp$$

i przechodnią. W takim razie, zgodnie ze znanym w logice matematycznej twierdzeniem o abstrakcji, można w zbiorze U wskazać klasy abstrakcji, czyli zbiory stanów

rzeczy nieodróżnialnych ze względu na relację N . Definicja pojęcia klasy abstrakcji ze względu na relację N brzmi następująco:

$$p \in [r_i]N, \text{ gdy } pNr_i,$$

czyli: stan rzeczy p należy do klasy abstrakcji relacji N wyróżnionej przez stan rzeczy r_i (ze zbioru U), gdy p pozostaje w relacji N do r_i . Intuicyjnie rzecz biorąc, zbiór $[r_i]N$ to zbiór stanów rzeczy równie obojętnych, co stan rzeczy r_i , czyli zbiór stanów rzeczy, z których żaden nie jest preferowany przed r_i , a przy tym takich, że r_i nie jest preferowany przed jakimkolwiek z nich.

Wiadomo z teorii mnogości, że relacja równościowa dzieli zbiór, na którym jest określona. W takim razie zbiór A wszystkich klas abstrakcji zsumowanych z sobą równy jest zbiorowi U , a ponadto — co dla nas w tej chwili ważniejsze — poszczególne klasy abstrakcji są ze sobą rozłączne, czyli nie mają wspólnych elementów; nie istnieje więc taki stan rzeczy p , który by był zarazem elementem dwóch różnych klas abstrakcji. Wniosek ten pozwala przyjąć nową definicję. Otóż preferowaniem globalnym nazywać będziemy relację P^* określoną następująco:

$$ZP^*Z', \text{ gdy } Z \in A \text{ i } Z' \in A \text{ i dla dowolnego } p \in Z, \text{ dla dowolnego } q \in Z': pPq.$$

Słowem, klasa abstrakcji Z jest preferowana globalnie przed klasą abstrakcji Z' , gdy każdy stan rzeczy ze zbioru Z jest preferowany przed wszystkimi stanami rzeczy ze zbioru Z' .

Łatwo zauważyć, że relacja preferowania globalnego P^* jest asymetryczna:

$$ZP^*Z' \rightarrow \sim (Z'P^*Z),$$

przechodnia:

$$(ZP^*Z') \cdot (Z'P^*Z'') \rightarrow ZP^*Z''$$

i spójna:

$$(ZP^*Z') + (Z'P^*Z)$$

(wśródzie tu Z , Z' i Z'' są dowolnymi zbiorami równocennych stanów rzeczy ze zbioru klas abstrakcji A). W takim więc razie relacja P^* porządkuje zbiór A , tzn. pozwala poszczególne klasy abstrakcji (zbiory równocennych stanów rzeczy) ustawić w pewien szereg — od zawierających najbardziej preferowane stany rzeczy (w sensie zwykłej relacji preferencji P) do zawierających najmniej preferowane stany rzeczy (w sensie tejże relacji P).

Zastanówmy się teraz, czy jest w systemie von Wrighta, scharakteryzowanym przez przyjęcie warunków asymetryczności i przechodniości dla relacji P , prawdą, że z dwóch stanów rzeczy p , $\sim p$ pierwszy jest bardziej preferowany od drugiego lub drugi od pierwszego. Innymi słowy rozważmy, czy prawdą jest, że

$$(pP\sim p) + (\sim pPp).$$

Załóżmy, że teza ta jest prawdziwa. Może ona być prawdziwa w trzech wypadkach: (A) kiedy oba jej człony są prawdziwe, (B) kiedy prawdziwy jest tylko pierwszy człon, (C) kiedy prawdziwy jest tylko drugi człon. Wypadek (A) jest z góry wykluczony — podstawmy bowiem $\sim p$ za q w postulatcie (a), a zobaczymy, że jeśli pierwszy człon tej alternatywy jest prawdą, to drugi jest fałszem. Załóżmy tedy, że ma miejsce przypadek (B). Otóż znowu podstawmy w (a) $\sim p$ za q . Ponieważ z założenia mamy $pP\sim p$, przeto uzyskujemy w ten sposób $\sim(\sim pPp)$. Z drugiej jednak strony założenie $pP\sim p$ prowadzi (po podstawieniu $\sim p$ za p) do wniosku, że $\sim pP\sim p$, to zaś jest równoważne tezie $\sim pPp$, sprzecznej z otrzymanym wyżej wnioskiem z postulatu (a). Załóżmy więc z kolei, że zachodzi wypadek (C), czyli że

prawdą jest teza, iż $\sim pPp$. Podstawiając w postulatcie (a) $\sim p$ za p i p za q otrzymujemy wobec prawdziwości powyższej tezy twierdzenie: $\sim (pP \sim p)$. Podstawmy jednak w tezie $\sim pPp \sim p$ za p , a otrzymamy: $\sim \sim pP \sim p$, co jest równoważne tezie $pP \sim p$, sprzecznej z otrzymanym wyżej wnioskiem z postulatu (a). Jak więc widać, założenie prawdziwości twierdzenia głoszącego, iż $pP \sim p$ lub $\sim pPp$ prowadzi do sprzeczności. Można w takim razie przyjąć, że

$$(pP \sim p) + (\sim pPp) + (pN \sim p).$$

Załóżmy jeszcze, że spełniony jest warunek¹:

$$(pN \sim p) \cdot (qN \sim q) \rightarrow pNq.$$

Czyli: jeśli stan rzeczy p jest równocenny ze stanem rzeczy nie- p , a stan rzeczy q jest równocenny ze stanem rzeczy nie- q , to p jest równocenne z q . Przyjęcie tego warunku wydaje się intuicyjne — wszak jeśli obojętne jest dla kogoś, czy zda egzamin, czy też nie, i jeśli obojętne jest dlań, czy otrzyma papierosy w kiosku, czy nie, to złożenie egzaminu ma dlań taką samą wartość jak kupno papierosów: nie preferuje ani złożenia egzaminu przed kupnem papierosów, ani kupna papierosów przed złożeniem egzaminu.

W świetle dotychczasowych ustaleń widać, że wśród klas abstrakcji (zbiorów równocennych stanów rzeczy) istnieje taki zbiór, którego elementami są stany rzeczy równocenne ze swymi negacjami, i że jest jeden tylko taki zbiór. Zbiór ten oznaczmy symbolem I i nazwijmy zbiorem neutralnych aksjologicznie stanów rzeczy. Tak więc stan rzeczy p należy do zbioru I , gdy $pN \sim p$, czyli gdy ani nie jest tak, że $pP \sim p$, ani nie jest tak, że $\sim pPp$.

Możemy teraz określić pojęcie dobra:

p jest dobry, gdy dla pewnej klasy Z ze zbioru A : $p \in Z$ oraz

$$ZP*I$$

i odpowiednio pojęcie zła:

p jest zły, gdy dla pewnej klasy Z ze zbioru A : $p \in Z$ oraz

$$IP*Z.$$

W uzupełnieniu tych definicji zauważmy, że:

1. Nie podlegają one zarzutowi, który wysunąłem wyżej pod adresem definicji von Wrighta. Bowiem, choć skazaniec może preferować śmierć przez powieszenie przed śmiercią w wyniku tortur (czyli przed nie-śmiercią przez powieszenie, jako że znajduje się on w sytuacji przymusowej i stoi przed tymi dwoma tylko alternatywami), to jednak nie staje się ona przez to dobrem w naszym sensie, bo przecież nie jest prawdą, że skazaniec preferuje śmierć przez powieszenie przed jakimkolwiek neutralnym dlań aksjologicznie stanem rzeczy. Nie stosuje się do naszej definicji także drugi zarzut, jaki wysunąłem przeciwko własnej modyfikacji definicji von Wrighta — nie przyjmujemy, że jeśli ktoś wybrał p przed q , to dla owego kogoś jest tak, że pPq . Takie rozumienie pojęcia preferencji byłoby nieintuicyjne: wyboru można istotnie dokonać wskutek posiadania autentycznych preferencji odpowiedniego rodzaju, ale też np. wskutek nacisku z zewnątrz, któremu nie można się przeciwstawić.

2. Definicje te pozwalają też na określenie pokrewnych pojęć aksjologicznych. Na przykład ideał możemy określić jako stan rzeczy należący do krańcowej klasy

¹ Na potrzebę wprowadzenia tego warunku wskazał mi J. Kmita.

abstrakcji w zbiorze A , czyli do klasy wyprzedzającej każdą inną pod względem P^* lub wyprzedzanej przez każdą inną pod tym względem. Z kolei p jest ideałem pozytywnym (negatywnym), gdy p jest ideałem i p jest dobrem (złem).

Doskonałym aksjologicznie (w sensie intencjonalnym) dla pewnego podmiotu można by nazwać człowieka, który zmierza do czynienia wyłącznie dobrych (w rozumieniu tego podmiotu) stanów rzeczy lub neutralnych (w rozumieniu owego podmiotu) stanów rzeczy, pewne zaś z owych dobrych stanów rzeczy są (dla owego podmiotu) ideałami pozytywnymi. Jeżeli ktoś jest nie tylko doskonały aksjologicznie w sensie intencjonalnym, ale i w sensie efektywnym, tj. czyni to wszystko, co zamierza, to jest on doskonały aksjologicznie po prostu (z punktu widzenia odnośnego podmiotu).

3. Wśród ideałów (a także ludzi doskonałych aksjologicznie) można by wyróżnić ideały realizowalne i nierealizowalne (w świetle wiedzy podmiotu, do którego wszystkie te pojęcia są zrelatywizowane). Te ostatnie wykazują daleko idące formalne podobieństwo do typów idealnych, jakimi operuje się w naukach empirycznych². W takim więc razie w pewnych przynajmniej przypadkach wartościowanie można by pojmować jako „procedurę” analogiczną do konkretyzacji typów idealnych³: stopień, w jakim określony stan rzeczy jest wartościowany, dałby się ustalić „odległością” od ideału pojętego jako typ idealny (tj. ideału nierealizowalnego). Być może dałoby się na tej drodze osiągnąć jakieś ciekawsze wyniki niż wyjaśnienie zbieżności etymologicznej słów „ideał”, „typ idealny” czy zwyczajowo terminologicznego usprawiedliwiającego np. powiedzenie, że miłość zawsze jest oparta na idealizacji.

4. Nie przesądzamy tu nic — wzorem von Wrighta — o merytorycznej „treści” relacji P , której charakterystyka jest wyłącznie formalna i pozwala relację tę interpretować nie tylko jako „lepszość moralną”, ale i jako „lepszość estetyczną”, czy jakąkolwiek inną.

Leszek Nowak

Лешек Новак

ДОБРО И ЗЛО С ТОЧКИ ЗРЕНИЯ ЛОГИКИ ПРЕФЕРЕНЦИИ*

Г. Г. фон Райт следующим образом определил понятие добра: положение вещей p является хорошим (для определенного субъекта) тогда, когда данный субъект положение вещей p предпочитает над положением вещей не- p . Это определение ведет к заключению, что например в случае, когда приговоренному к смертной казни предоставляется возможность выбора вида смерти или через повешение, или в эффекте пыток и если избирает он первый вид смерти, тогда смерть через прवेशение является для него добром.

В статье делается попытка устранения этой консеквенции, но так как ведет она к дефиниции, которая подвергнута на другие недостатки, автор — используя терминологию теории Райта — предлагает иное определение добра и зла.

² Eksplikację pojęcia idealizacji i typu idealnego przedstawiam w pracy *U podstaw Marksowskiej metodologii nauk*, Warszawa 1971.

³ Pojęcie konkretyzacji omawiam w wymienionej w przypisie 2 pracy.

* В связи с книгой: G. H. von Wright, *The Logic of Preference*, Edinburgh 1963.

Leszek Nowak

GOOD AND EVIL IN THE LOGIC OF PREFERENCE *

G. H. von Wright's definition of the concept of good reads: the state of a thing p is good (for a definite subject) when it is preferred by that subject before the state of affairs non- p . This definition implies that when a man sentenced to death has no other choice except between death on the gallows or after tortures and when he chooses the former kind of death, then death by hanging is good for him. A certain way of escaping this consequence is outlined but since it leads to a definition vulnerable to other objections a different definition of the concepts of good and evil in the term of G. H. von Wright's theory is suggested.

* Cf. G. H. von Wright, *The Logic of Preference*, Edinburgh 1963.