# R E C E N Z J E

ROCZNIKI FILOZOFICZNE
Tom LXVI, numer 2 – 2018

## WZNOWIENIE FILOZOFICZNEGO WPROWADZENIA DO TEORII DECYZJI

Martin PETERSON, *An Introduction to Decision Theory*. (Cambridge Introductions to Philosophy). Cambridge: Cambridge University Press, 2017, 339 ss. ISBN 978-1-107-15159-8 (Hardback). ISBN 978-1-316-60620-9 (Paperback).

DOI: http://dx.doi.org/10.18290/rf.2018.66.2-11

Teoria decyzji jest już obecnie dyscypliną o dobrze ugruntowanej pozycji. Zewnętrznym świadectwem tej pozycji są chociażby Nagrody Nobla — przede wszystkim w dziedzinie ekonomii — które kilkukrotnie już przyznano za prace z zakresu teorii decyzji (laureatami są m.in. John Nash, Kenneth Arrow, John Hersanyi, Amartya Sen, Daniel Kahneman). Coraz większy jest też wpływ tej teorii na filozofię, widoczny szczególnie wyraźnie w etyce oraz filozofii społecznej i politycznej. W tym kontekście z radością należy przyjąć drugie wydanie filozoficznego wprowadzenia do teorii decyzji (*An Introduction to Decision Theory*) autorstwa Martina Petersona, amerykańskiego filozofa i etyka z Texas A&M University.

Recenzowana praca składa się z czternastu rozdziałów. Choć Autor nie grupuje ich w żadne części, można w przybliżeniu podzielić je na cztery działy. Pierwszy zawiera rozdziały wprowadzające podstawowe terminy, dystynkcje oraz narzędzia teorii decyzji. Drugi dotyczy teorii decyzji indywidualnych, trzeci – teorii gier, a czwarty – teorii wyboru społecznego.

W otwierającym podręcznik rozdziale pierwszym Peterson podaje rozróżnienie na normatywną (szukającą wzorców racjonalnego działania) oraz deskryptywną (opisującą faktyczne schematy działania) teorię decyzji. Poza ostatnim rozdziałem, prezentującym teorię perspektywy, książka poświęcona jest wykładowi teorii normatywnej. Autor wymienia wspomniane wyżej subdyscypliny teorii decyzji, odróżnia decyzje racjonalne od właściwych oraz podaje krótką historię dziedziny. W rozdziale drugim umieszczone zostały wyjaśnienia takich podstawowych pojęć, jak: stany (*states*), wypłaty (*outcomes*) oraz działania (*actes*). Większość wprowadzonej tu terminologii przewija się przez cały podręcznik.

Do rozdziałów o charakterze wprowadzającym zaliczyłbym także rozdziały szósty i siódmy. Poświęcone są one wykładowi teorii prawdopodobieństwa, a więc jednego z głównych narzędzi formalnych wykorzystywanych w teorii decyzji. Pierwszy z nich (szósty) zawiera przystępną prezentację podstawowych matematycznych własności prawdopodobieństwa, podczas gdy drugi (siódmy) dostarcza filozoficznych interpretacji tego matematycznego obiektu. Rozdziały są dobrze napisane, a materiał odpowiednio dobrany, dlatego mogą one służyć jako niezależne od reszty podręcznika krótkie filozoficzne wprowadzenie do teorii prawdopodobieństwa.

Druga z wyróżnionych przeze mnie roboczo części recenzowanej książki dotyczy teorii decyzji indywidualnych, tzn. podejmowanych przez jednostkę w sytuacjach niezależnych od działań innych jednostek. Na tą część składałyby się rozdziały od trzeciego do dziesiątego oraz zamykający książkę rozdział czternasty, z wyjątkiem wspomnianych wyżej rozdziałów szóstego i siódmego.

Rozdział trzeci poświęcony jest prezentacji decyzji podejmowanych w warunkach niepewności, a więc decyzji, w których wypadku podmiot działania nie zna prawdopodobieństw zajścia poszczególnych możliwych stanów świata. Poza oczywistym i uniwersalnym kryterium dominacji zwykle wymienia się przynajmniej cztery kryteria, które uważane są za wzorce racjonalnych działań we wspomnianych sytuacjach (reguły: maksiminu, optymizmu-pesymizmu, minimaksu żalu oraz zasadę niewystarczającej racji). Peterson omawia wszystkie te reguły, a ponadto opisuje pewne inne propozycje.

Rozdział czwarty dotyczy decyzji podejmowanych w warunkach ryzyka, w których przypadku prawdopodobieństwa zajścia możliwych stanów świata są podmiotowi działania znane. W przypadku tej klasy sytuacji decyzyjnych wskazywana jest przede wszystkim jedna reguła wyznaczająca działania racjonalne – zasada maksymalizacji wartości oczekiwanej. Rozdział czwarty jest poświęcony jej aksjomatycznej postaci oraz trudnościom wynikającym z pewnych paradoksów (Allaisa, Ellsberga, gry petersburskiej i in.).

Pozostałe rozdziały tej części dotyczą von Neumana-Morgensterna teorii przedziałowej skali użyteczności (rozdział piąty), bayesianizmu w teorii decyzji (rozdział ósmy), wyrosłego wskutek dyskusji wokół problemu Newcomba podziału na ewidencjalną i przyczynową teorię decyzji (rozdział dziewiąty) oraz trzech postaci niechęci do ryzyka (rozdział dziesiąty). W ostatnim rozdziale Peterson opisał nagrodzoną Nagrodą Nobla teorię perspektywy Daniela Kahnemnana i Amosa Tverskiego, która stanowi przykład deskrypcyjnej teorii decyzji i opisuje faktyczne procesy rządzące podejmowaniem decyzji przez ludzi.

Część poświęcona teorii gier, a więc sytuacjom decyzji, w których możliwe wyniki zależą częściowo od działań innych jednostek, składa się z rozdziałów jedenastego i dwunastego. Omówione są tam podstawowe rodzaje gier, zasada dominacji, punkty równowagi Nasha oraz odróżnienie strategii czystych i mieszanych. Sporo uwagi poświęcono dwuosobowym grom o sumie zerowej oraz różnym przykładom gier o sumie niezerowej. Autor wskazał także implikacje omawianych pojęć w zakresie zagadnień ewolucji oraz problemów etycznych.

Ostatnia część, dotycząca teorii wyboru społecznego, składa się z zaledwie jednego (trzynastego) rozdziału. Jego przedmiotem są sposoby racjonalnego ustalania preferencji społeczności jako całości na podstawie znanych preferencji wszystkich członków tej społeczności. Szczególnie ciekawe filozoficznie są rozważania dotyczące twierdzenia Arrowa, stwierdzającego, że nie istnieją sposoby ustalania preferencji społecznych posiadające pewne — łączone zwykle z demokracją — właściwości (niezależność od nieistotnych alternatyw, własność Pareto, brak dyktatury, uniwersalność). Nie mniej zajmujące są rozważania dotyczące twierdzenia A. Sena o liberalizmie i jego dyskusji z Robertem Nozickiem.

Główny tekst uzupełniony został numerowanymi ramkami, które zawierają zwykle fragmenty bardziej zaawansowane technicznie, głównie dowody twierdzeń. Ponadto każdy rozdział zakończony jest zadaniami z dołączonymi do nich odpowiedziami. Jest to rozwiązanie bardzo pomocne w dydaktycznym wykorzystaniu podręcznika. Poza tym na końcu zamieszczony został słowniczek podstawowych pojęć, co także ułatwia korzystanie z książki.

Należy jeszcze wspomnieć o różnicy między wydaniem pierwszym a recenzowanym wznowieniem. Poza dodaniem kilkudziesięciu zadań oraz dołączeniem kilku nowych przykładów i odwołań do najświeższej literatury Autor zdecydował się także połączyć dwa oddzielne w wydaniu pierwszym rozdziały w jeden (ósmy), poświęcony bayesianizmowi, a także dodać rozdział poświęcony niechęci do ryzyka. Dokładniejszy opis wprowadzonych zmian znaleźć można w przedmowie do wydania drugiego.

Dużym plusem recenzowanej pracy jest przystępny styl prezentacji licznych terminów formalnych. Uważam za dobrą strategię przedstawianie najpierw intuicji stojących za pewnymi pojęciami technicznymi, a dopiero później formalną ekspozycję tych pojęć. Dobór materiału, duża liczba przykładów oraz liczne odniesienia do zagadnień *stricte* filozoficznych stanowią niewątpliwe zalety podręcznika Petersona.

Niestety książka ma też poważną wadę, a mianowicie sporą liczbę błędów. Pojawiają się one zwłaszcza w formalnych fragmentach wykładu, a także w zadaniach oraz odpowiedziach do tychże. Błędy te mają dwojaką naturę. Po pierwsze, pojawiają się drobne potknięcia, które można potraktować jak literówki (brak nawiasu, niepoprawny znak arytmetyczny, brak znaku pochodnej, pomylony indeks itp.). Autor opublikował na swojej stronie internetowej erratę (http://www.martinpeterson.org/; zakładka *Decision theory*), jednak liście skorygowanych tam potknięć daleko do kompletności (co gorsza, w drugim wydaniu pojawiają się niekiedy błędy w miejscach, w których w wydaniu pierwszym błędów nie było). Po drugie, niekiedy pojawiają się jednak nieco poważniejsze usterki. Obecna recenzja nie jest miejscem na szczegółową prezentację takich pomyłek. Podam trzy przykłady, które umożliwią Czytelnikowi wyrobienie sobie zdania na temat ich natury i wagi.

Pierwszy przykład dotyczy poświęconego wypłatom paragrafu 2.2, a dokładnie końcowego fragmentu tego paragrafu, prezentującego rozróżnienie trzech rodzajów skal, które mogą być użyte do pomiaru wartości wspomnianych wypłat. Omawiane skale to skala porządkowa (*ordinal*), przedziałowa (*interval*) oraz stosunkowa (*ratio*). Na s. 27 Autor pisze, że funkcja $f$ jest skalą interwałową wtedy i tylko wtedy, gdy spełnione są dwa warunki: (1) $f(x) \geq f(y)$ wtedy i tylko wtedy, gdy $x \geq y$,[1] tzn. wypłaty wyżej cenione mają przypisane wyższe wartości liczbowe na skali, oraz (2) dla każdej innej funkcji $f'$ spełniającej warunek (1) istnieją jakieś stałe $k$ i $m$ takie, że $f'(x) = k \cdot f(x) + m$. Warunek (2) (w tekście oznaczony numerem 4) wymaga, aby każda transformacja oryginalnej skali interwałowej $f$ zachowująca jej porządek była transformacją liniową. Taki warunek nie może być spełniony dla żadnej funkcji $f$, zawsze bowiem można znaleźć inną funkcję zachowującą porządek oryginalnej funkcji, niebędącą jednak liniową transformacją tej pierwszej. Jest to dość oczywiste stwierdzenie, ale aby je unaocznić (bez dowodu), podam prosty przykład. Załóżmy, że $f(x) = x$, a $f'(x) = \sqrt{x}$ oraz, że $x > 0$. Druga funkcja ($f'$) spełnia warunek (1), nie jest jednak liniową transformacją funkcji $f$. Ponieważ warunek (2) nigdy nie jest spełniony, opisane przez Autora pojęcie skali interwałowej musi być pojęciem pustym! To samo dotyczy podanego w tekście opisu skali stosunkowej.

Przedstawiony przypadek, choć nie jest zwykłą literówką, to mimo wszystko pozostaje niezbyt groźnym, gdyż dość oczywistym potknięciem. Z innych fragmentów paragrafu 2.2 wynika bowiem, że wspomniane liniowe transformacje oryginalnej skali to nie wszystkie możliwe przekształcenia zachowujące porządek, ale tylko tzw. transformacje dopuszczalne. Przez te ostatnie rozumie Autor przekształcenia zachowujące nie tylko porządek, ale również zawarte w oryginalnej skali informacje na temat względnych różnic (przedziałów) między mierzonymi wypłatami. Innymi słowy, dokonując liniowej transformacji jakiejś skali interwałowej, otrzymamy inną skalę interwałową równoważną tej pierwszej (przykładowo informacja na temat temperatury powietrza podana na skali Celsjusza jest równoważna informacji przedstawionej na skali Fahrenheita; każda z tych skal jest liniową transformacją drugiej).

Nieco poważniej wygląda drugi przykład, który dotyczy rozdziału ósmego, poświęconego bayesiańskiej teorii decyzji. W paragrafie 8.2 Peterson omawia pragmatyczny argument na rzecz przechodniości relacji preferencji, zwany argumentem z oskubania (*money-pump argument*). Celem tego rozumowania jest wykazanie, że nieprzechodnie, a dokładniej kołowate preferencje mogą być wykorzystane przeciw osobie żywiącej takie preferencje. Załóżmy, że ktoś (Euzebiusz) ceni wyżej $x$ niż $y$, $y$ ceni wyżej niż $z$, a jednocześnie $z$ ceni wyżej niż $x$ (takie preferencje są kołowate, a więc nieprzechodnie). Załóżmy ponadto, że wspomniany osobnik posiada $y$. Ponieważ ceni on $x$ wyżej niż $y$, na pewno znajdzie się jakaś niewielka suma pieniędzy $\varepsilon_1$, którą będzie gotów zapłacić za wymianę posiadanego przez siebie $y$-a na $x$-a. Następnie, ponieważ Euzebiusz ceni $z$ wyżej niż $x$, znajdzie się znowu jakaś niewielka suma pieniędzy $\varepsilon_2$, którą będzie gotów zapłacić za wymianę $x$-a na $z$-a. Wreszcie, ponieważ Euzebiusz ceni $y$ wyżej niż $z$, znów będzie gotów zapłacić jakąś niewielką sumę $\varepsilon_3$ za

[1] Już tutaj sformułowanie nie jest precyzyjne, gdyż wypłaty $x$ i $y$ nie muszą być wcale liczbami, a więc relacja większy lub równy nie musi się do nich stosować.

wymianę $z$-a na $y$-a. Po tych trzech transakcjach Euzebiusz jest w posiadaniu $y$-a, tak jak był w jego posiadaniu przed przystąpieniem do serii wymian, jednocześnie jednak jest uboższy o $\varepsilon_1 + \varepsilon_2 + \varepsilon_3$. Po odpowiednio dużej liczbie transakcji opisanego rodzaju Euzebiusz zostanie bankrutem (będzie „oskubany" z pieniędzy), nie zyskując w zamian nic.

Poza opisowym przedstawieniem powyższego argumentu Autor podaje — w ramce 8.1 (s. 181–182) — jego wersję formalną. Główną tezą jest twierdzenie 8.1, stwierdzające sprzeczność podanych wcześniej dwóch przesłanek oraz pewnego dodatkowego założenia T. Zawierają one dość naturalne obserwacje dotyczące predykatu dopuszczalności, a ich dokładna treść nie jest tutaj istotna. Dowód nie wprost tego twierdzenia zaczyna się od założenia, że relacja preferencji jest przechodnia (sic!), z którego to założenia, w połączeniu z wykorzystaniem wspomnianych przesłanek oraz założenia T, wyciąga się sprzeczność. Wspomniany dowód[2] wykazuje zatem nie tyle sprzeczność przesłanek oraz założenia T, co raczej ich sprzeczność z przechodniością relacji preferencji. To jednak jest niezrozumiałe, a nawet zaskakujące w świetle faktu, że wzmiankowany paragraf jest poświęcony obronie przechodniości relacji preferencji. Naturalne byłoby raczej oczekiwanie dowodu sprzeczności wspomnianych przesłanek i złożeń z nieprzechodniością lub przynajmniej kołowatością preferencji.

Trzeci przykład zawiera trudność jeszcze poważniejszą. Dotyczy dowodu twierdzenia minimaksowego (*minimax theorem*), przedstawionego w ramce 11.2 (s. 256–258), zamieszczonej pod koniec poświęconego teorii gier rozdziału jedenastego. Zgodnie ze wzmiankowanym twierdzeniem każda dwuosobowa gra o sumie zerowej ma rozwiązanie, a więc parę strategii znajdujących się w równowadze. Para strategii znajduje się w równowadze wtedy i tylko wtedy, gdy żaden z graczy nie poprawi swojego wyniku, jednostronnie zmieniając strategię, żadnemu graczowi nie opłaca się jednostronna zmiana.

W dowodzie twierdzenia minimaksowego Peterson wprowadza pojęcie standardowej postaci dwuosobowej gry o sumie zerowej. Jest to gra, której macierz wypłat wygląda następująco:

| | C1 | C2 |
|---|---|---|
| R1 | a | d |
| R2 | c | b |

przy tym wszystkie wielkości w tabeli są dodatnie, a ponadto $a > c$, $b > c$ oraz $b > d$. Autor zaznacza, że przy takich założeniach nie istnieje para czystych (*pure*) strategii znajdujących się w równowadze.

[2] W dowodzie pojawiają się ponadto drobniejsze potknięcia. W drugim i trzecim jego wierszu jest „wybór pomiędzy $x$ w $t_1$ i pomiędzy $x$ i $y$ w $t_2$", a powinno być „wybór pomiędzy $x$ i $y$ w $t_1$ i pomiędzy $y$ i $z$ w $t_2$". Pomyłka ta nie pojawia się w erracie zamieszczonej przez Petersona na jego stronie internetowej.

Łatwo przekonać się, że powyższy sąd jest fałszywy. W tym celu wystarczy wybrać jakiekolwiek wartości spełniające następujący warunek: $b > d > a > c$. Gwoli przykładu niech $a = 2$, $b = 4$, $c = 1$ oraz $d = 3$, co da w wyniku następującą macierz wypłat

| | C1 | C2 |
|---|---|---|
| R1 | 2 | 3 |
| R2 | 1 | 4 |

Para strategii czystych (R1, C1) pozostaje w równowadze. Graczowi R nie opłaca się zmiana na R2, gdyż zamiast 2 otrzymałby 1. Graczowi C także nie opłaca się zmiana na C2, ponieważ zamiast stracić 2, straciłby 3 (omawiana gra jest grą o sumie zerowej, a więc zysk jednego gracza jest równy stracie drugiego).

To jednak nie koniec trudności obecnych w omawianym dowodzie. Ponieważ Autor uważa (błędnie), że nie ma pary strategii czystych pozostających w równowadze, przechodzi do wskazania pary strategii mieszanych (*mixed*). W tym celu dokonuje serii przekształceń algebraicznych (w których też pojawiają się błędy, wskazane wszelako w erracie udostępnionej przez Autora na jego stronie internetowej), prowadzących do wniosku, że gracz R powinien wybrać strategię R1 z prawdopodobieństwem $p = M/K$, przy czym $M = |c\ -b|$, a $K = |a - d - c + b|$. Jeśli za wartości zmiennych $a$, $b$, $c$, $d$ przyjmiemy te zaproponowane w poprzednim akapicie, wówczas prawdopodobieństwo $p = 1{,}5$ (sic!).

Twierdzenie minimaksowe zostało udowodnione przez Johna von Neumanna, można zatem być spokojnym o jego prawdziwość. Dowód jednak zamieszczony w recenzowanym podręczniku wymaga poważnej korekty.

Jest dość niepokojące, że tak szanowane wydawnictwo jak Cambridge University Press wydaje wznowienie nie tylko powielające większość potknięć wydania pierwszego, ale jeszcze dodatkowo zawierające nowe błędy, również czysto edytorskie[3]. Jest to tym bardziej dziwne, gdyż recenzenci pierwszego wydania zwracali uwagę na liczne jego błędy (por. recenzja Bena Egglestone'a w *Notre Dame Philosophical Reviews. An Electronic Journal*; https://ndpr.nd.edu/news/an-introduction-to-decision-theory/).

W świetle powyższych uwag moja ocena książki Martina Petersona pozostaje ambiwalentna. Z jednej strony nie mogę jej polecić studentom, zwłaszcza do samodzielnej lektury. Nauka z podręcznika zawierającego błędy może okazać się frustrująca. Nigdy nie ma bowiem pewności, czy ewentualne problemy ze zrozumieniem fragmentów wynikają z zawiłości materii, czy z omyłek w tekście. Z drugiej strony książka Petersona zawiera przystępną (poza wspomnianymi błędami) prezentację bogatych treści z zakresu teorii decyzji. Na rynku wydawniczym, również anglo-

[3] Przykładowo w nowym wydaniu, w dodanym rozdziale poświęconym niechęci do ryzyka, na s. 224 pojawia się cytat z pracy Matthew Rabina z 2000 r. Nazwisko Rabina nie pojawia się jednak ani w indeksie końcowym, ani w bibliografii.

językowym, nie ma zbyt wielu pozycji konkurencyjnych[4]. Recenzowany podręcznik może być dobrym wyborem pod warunkiem wyczulonej na błędy lektury, szczególnie uważnej w przypadku fragmentów prezentujących techniczne aspekty omawianych zagadnień.

*Piotr Lipski*
*Katedra Teorii Poznania*
*na Wydziale Filozofii KUL*
*e-mail: piotr.lipski@kul.pl*

[4] Przykładem konkurencyjnej pozycji może być Michael RESNIK, *Choices. An Introduction to Decision Theory* (Minneapolis, London: University of Minnesota Press, 1987).