MODELE NIEEGOISTYCZNYCH PREFERENCJI: WPROWADZENIE I NAJNOWSZE BADANIA

Anna Kosior*

Streszczenie: *Dynamiczny rozwój ekonomii eksperymentalnej, w szczególności jej nurtu związanego z teorią gier, przyczynił się do zakwestionowania podstawowych założeń, na których opiera się klasyczna teoria gier – założenia o racjonalności oraz o egoizmie graczy. W rezultacie zaczęto modyfikować powyższe założenia i tworzyć „bliższe rzeczywistości" modele postępowania ludzi w sytuacjach, które dają się opisać jako gry. Artykuł stanowi krótki przegląd tych modeli, które rezygnują z założenia o egoistycznych preferencjach. W artykule analizowane są przede wszystkim modele dystrybucyjne, w tym modele awersji do nierównych wyników podziału oraz model preferencji quasi-maksiminowych. Prezentowane są także rezultaty eksperymentów projektowanych w celu testowania konkurencyjnych modeli dystrybucyjnych oraz wyniki eksperymentów, które badają znaczenie wzajemności opartej na ocenie intencji jako czynnika motywującego zachowania ludzi.*

Słowa kluczowe: *gry eksperymentalne, behawioralna teoria gier, modele nieegoistycznych preferencji, awersja do niesprawiedliwych wyników podziału, wzajemność oparta na intencjach.*

MODELS OF SOCIAL PREFERENCES: INTRODUCTION AND RECENT RESEARCH

Abstract: *: The dynamic development of experimental economics, in particular of experimental work in game theory, has contributed to the questioning of the basic assumptions of classical game theory: the rationality and self-interest assumption. In response, those assumptions are being modified and more realistic models of human behavior in situations which can be described as games are being developed. The article provides a short overview of the models that abandon the self-interest assumption. An account is given mostly of the distributional models (i.a. inequity aversion models,*

* Anna Kosior, Narodowy Bank Polski, ul. Świętokrzyska 11/21, 00-919 Warszawa, fax 022 826 99 35, e-mail: anna.kosior@mail.nbp.pl

model of quasi-maximin preferences). The results of the experiments designed to test the rival distributional models of social preferences are also presented in the article, as well as of the experiments that assess to the role played by intention-based reciprocity in motivating human behavior

Keywords: *experimental games, behavioral game theory, other-regarding preferences, social preferences models, inequity aversion, intention-based reciprocity.*

Wprowadzenie

Klasyczna teoria gier opiera się na wysoce wyidealizowanych założeniach behawioralnych. W myśl tej teorii decydenci są racjonalnymi egoistami. Podejmują działania optymalne z punktu widzenia założonego celu, jakim jest osiągnięcie najbardziej preferowanego wyniku (racjonalność). Najbardziej preferowany jest zaś ten wynik, który pozwala na maksymalizację własnej użyteczności. Założenie o racjonalnym i egoistycznym zachowaniu decydentów najczęściej ujmowane jest w kategoriach maksymalizacji indywidualnej funkcji użyteczności, która zależy wyłącznie od materialnych wypłat danego gracza, całkowicie pomija natomiast materialne wypłaty innych graczy (egoizm).

Jak łatwo jednak zauważyć, faktyczni decydenci rzadko potrafią jednoznacznie określić wszystkie czynniki decyzyjne, ich możliwości obliczeniowe są ograniczone, a zachowania nie w każdej sytuacji można określić jako racjonalne (Malawski i inni 2004: 47). Przewidywania odnośnie zachowań graczy wywodzone z matematycznych modeli teorii gier mogą więc znacznie odbiegać od tego, jak ludzie będą się rzeczywiście zachowywać w sytuacjach, których struktura interakcyjna jest taka sama jak struktura analizowanych gier. W ostatnich latach bardzo dynamicznie rośnie liczba eksperymentów, których celem jest testowanie tych przewidywań. W tym kontekście mówi się nawet o powstaniu nowej dyscypliny, jaką jest behawioralna teoria gier. Teoria ta rozszerza i modyfikuje behawioralne założenia klasycznej teorii gier, wprowadzając do analiz takie czynniki, jak: emocje, moralność, błędy poznawcze i obliczeniowe, ograniczone możliwości tworzenia planów działania i budowania prognoz, niepewność co do racjonalności pozostałych graczy, czy też możliwość uczenia się przez uczestników gry (Camerer 2003: 3).

Jednym z istotnych obszarów zainteresowań behawioralnej teorii gier jest założenie o egoistycznej naturze decydentów. Założenie to poddawane było intensywnym testom z wykorzystaniem licznych eksperymentów ekonomicznych[1]. Wyniki tych eksperymen-

[1] Systematyczny przegląd eksperymentów, które dały podstawę do zakwestionowania założeń teorii gier o racjonalności i egoizmie graczy, znaleźć można w książce C.F. Camerera (2003) oraz w pracy E. Fehra i K. Schmidta (2006). W języku polskim dostępne są artykuły S. Czarnika (2007) oraz T. Kopczewskiego i M. Malawskiego (2007).

tów sugerują, że znaczna część osób dba nie tylko o swoje własne materialne wypłaty, ale także o sprawiedliwość podziału i o sposób, w jaki traktowani są przez innych graczy[2]. Są oni skłonni do ponoszenia pewnych kosztów w celu zmiany podziału, który uznają za niesprawiedliwy. Są także skłonni do nagradzania osób, które wykazują chęć do współpracy i karania tych, którzy wyłamują się ze wspólnych przedsięwzięć. Regularne odchylenia faktycznych zachowań znacznej części uczestników eksperymentów od zachowań przewidywanych na podstawie założenia o egoistycznych preferencjach graczy skłoniły niektórych teoretyków do modyfikacji tego założenia czy wręcz całkowitej rezygnacji z niego. W rezultacie powstały ***modele nieegoistycznych preferencji***[3]. Modele te zachowują założenie klasycznej teorii gier o racjonalności decydentów, dopuszczając jednocześnie, by indywidualne funkcje użyteczności zależały także od materialnych wypłat otrzymywanych przez innych graczy. W artykule omówione zostaną najbardziej popularne modele tego typu. Zaprezentowane zostaną także wyniki badań eksperymentalnych poświęconych testowaniu różnych modeli nieegoistycznych preferencji[4].

---

[2] Zob. przykładowo analizę wyników licznych eksperymentów opartych na grze *Ultimatum* przeprowadzoną przez C.F. Camerera (2003: 48-83) lub omówienie wyników eksperymentów, które dały bodziec do tworzenia modeli nieegoistycznych preferencji, w pracy E. Fehra i K. Schmidta (2006).

[3] W literaturze przedmiotu w chwili obecnej brak jest jednoznacznej terminologii dla określenia tego typu modeli. W pracach anglojęzycznych poświęconych tej problematyce pojawiają się określenia takie jak *„social preferences"*, *„other-regarding preferences/behaviour"* czy też *„interdependant preferences"*, przy czym nie przez wszystkich autorów używane są one w ten sam sposób. Co więcej, niektórzy badacze wykazują niekonsekwencję w stosowaniu poszczególnych terminów. Zdarza się, że w różnych swoich pracach temu samemu terminowi nadają oni odmienne znaczenie [Przykładowo w jednym z artykułów, którego współautorem jest E. Fehr, stwierdza się, że dana osoba *„ma preferencje społeczne* (ang. social preferences), *jeżeli dba nie tylko o materialne zasoby, które przypadają jej samej, ale również o materialne zasoby innych, istotnych punktu widzenia tej osoby graczy"* (Fehr, Fischbacher 2002). Tak pojmowane *„preferencje społeczne"* są tożsame z tym, co w niniejszej pracy określa się jako *„preferencje nieegoistyczne"*. Z kolei w innej pracy, której współautorem jest także E. Fehr, preferencje społeczne są tylko jednym z wielu typów preferencji nieegoistycznych (Fehr, Schmidt 2006)]. Problem pojawia się także przy tłumaczeniu tych terminów na język polski. Najczęściej stosowane w literaturze anglojęzycznej określenie *„social preferences"* bezpośrednio przetłumaczone na język polski oznacza *„preferencje społeczne"*. Termin *„preferencje społeczne"* w literaturze poświęconej teorii racjonalnego wyboru tradycyjnie używany jest jednak dla określenia preferencji stanowiących wynik agregacji indywidualnych preferencji członków jakiejś grupy. Używanie tego terminu także dla określenia preferencji indywidualnych reprezentowanych za pomocą funkcji użyteczności, która oprócz wypłaty pieniężnej danego gracza uwzględnia dodatkowo wypłaty innych graczy, mogłoby rodzić różne niejasności i nieporozumienia. Dlatego też w niniejszej pracy dla określenia modeli z tego typu preferencjami będziemy się posługiwać terminem *„modele nieegoistycznych preferencji"*. Wymaga to jednak pewnego dodatkowego zastrzeżenia. Modele te nie wykluczają bowiem możliwości, że gracze mają preferencje egoistyczne. Jak zobaczymy w dalszej części pracy, model preferencji egoistycznych jest „zagnieżdżony" w większości modeli preferencji nieegoistycznych. Są to więc modele bardziej ogólne, w których dla pewnych wartości parametrów funkcje użyteczności są tożsame z funkcjami użyteczności standardowego modelu z preferencjami egoistycznymi.

[4] Artykuł ten oparty jest na pracy magisterskiej przygotowywanej przez autorkę w Instytucie Socjologii UW. Ze względu na ograniczenia objętościowe w artykule przedstawione zostaną tylko wybrane eksperymenty testujące modele nieegoistycznych preferencji graczy – przede wszystkim te najbardziej spopularyzowane w anglojęzycznej literaturze przedmiotu. Pewne generalizacje zawarte w każdym z podrozdziałów poświęconych tym eksperymentom formułowane są jednak w oparciu o szerszy przegląd wyników badań dokonany w wyżej wymienionej pracy. Czytelnik może znaleźć więcej informacji na temat tego typu eksperymentów w pracy E. Fehra i K. Schmidta (2006).

1. Dystrybucyjne modele nieegoistycznych preferencji graczy

W literaturze poświęconej modelom nieegoistycznych preferencji wyróżnia się najczęściej dwa typy takich modeli (np. Bolton, Ockenfels 2005: 958): 1) ***modele dystrybucyjne***, w których indywidualne funkcje użyteczności zależą tylko od końcowej alokacji materialnych wypłat w grze oraz 2) ***modele wzajemności***, w których preferencje zależą dodatkowo albo od tego, jak doszło do powstania danej alokacji (czyli *de facto* od tego, jak zachowywali się inni gracze), albo od tego, z jakim graczem dana gra była rozgrywana (zob. *rozdział 5* artykułu).

Wśród modeli dystrybucyjnych największą popularnością cieszą się ***modele awersji do niesprawiedliwych wyników gry*** (Bolton 1991[5]; Fehr, Schmidt 1999; Bolton, Ockenfels 2000) oraz ***model preferencji quasi-maksiminowych*** Charnessa i Rabina (2002).

W modelach awersji do niesprawiedliwych wyników gry sprawiedliwość zachowania graczy oceniana jest przez pryzmat *równości* wypłat pieniężnych w grze. Modele te przyjmują, że na użyteczności części decydentów negatywnie wpływa fakt, że końcowa alokacja materialnych wypłat w grze może być alokacją nierówną. Osoby te są gotowe zrezygnować z części przypadającego im dobra, by podział stał się bardziej równy. Ważną własnością obu modeli awersji do niesprawiedliwych wyników gry jest fakt, że dopuszczają one, by część graczy miała preferencje *stricte* egoistyczne. Interakcje między rozkładem heterogenicznych preferencji w populacji graczy, a kontekstem strategicznym[6] gry pozwalają wyjaśniać, dlaczego w niektórych eksperymentach dominować mogą zachowania egoistyczne, w innych zaś dominują rozwiązania egalitarne.

---

[5] W modelu Boltona (1991), który w tym artykule uznawany jest za pewien szczególny przypadek modeli awersji do niesprawiedliwych wyników gry, dany gracz odczuwa awersję do nierówności tylko wówczas, gdy otrzymuje on relatywnie mniej niż inny gracz. W przypadku pozostałych modeli zaliczanych do tej klasy modeli dystrybucyjnych na użyteczność gracza negatywnie wpływać może zarówno to, że dostaje on mniej niż inni, jak i to, że jego materialna wypłata jest większa niż wypłaty innych graczy. Z tego powodu np. E. Fehr i K. Schmidt (2006) traktują model Boltona jako przykład odrębnego typu modeli dystrybucyjnych, mianowicie *modeli zazdrości*.

[6] Przykładowo, w eksperymentach opartych na grze *Ultimatum*, w której jeden z graczy proponuje podział pewnej kwoty między siebie i drugiego gracza, a drugi gracz może ten podział zaakceptować bądź odrzucić (przy czym akceptacja jest równoważna z wprowadzeniem podziału w życie, natomiast brak akceptacji oznacza, że gracze nie otrzymują nic), zaobserwowano, iż modalne i mediany ofert w tej grze wynoszą zazwyczaj 40-50%, a średnie oferty – 30-40% dzielonej kwoty. Ponadto niemalże wcale nie pojawiają się oferty wynoszące 0%, 1-10% czy powyżej 51% sumy do podziału (Camerer 2003: 43-83). Wyniki te są niezgodne z przewidywaniami formułowanymi na podstawie założenia o egoistycznych preferencjach graczy. Gdy jednak zmienimy kontekst strategiczny, wprowadzając do gry kilka osób, które osobno decydują, czy zaakceptować, czy odrzucić zaproponowany podział, po kilku rundach wyniki gry zbiegną do wyników przewidywanych przez klasyczną teorię gier nawet wówczas, gdy gracze będą mieli nieegoistyczne preferencje (Fehr, Schmidt 1999: 825-831). Więcej na temat interakcji między kontekstem strategicznym gry a rozkładem preferencji w populacji graczy można przeczytać w pracy E. Fehra i K. Schmidta (2006).

W modelu Fehra i Schmidta funkcja użyteczności gracza *i* dana jest następującym wzorem (Fehr, Schmidt 1999):

$$U_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\} \qquad [1]$$

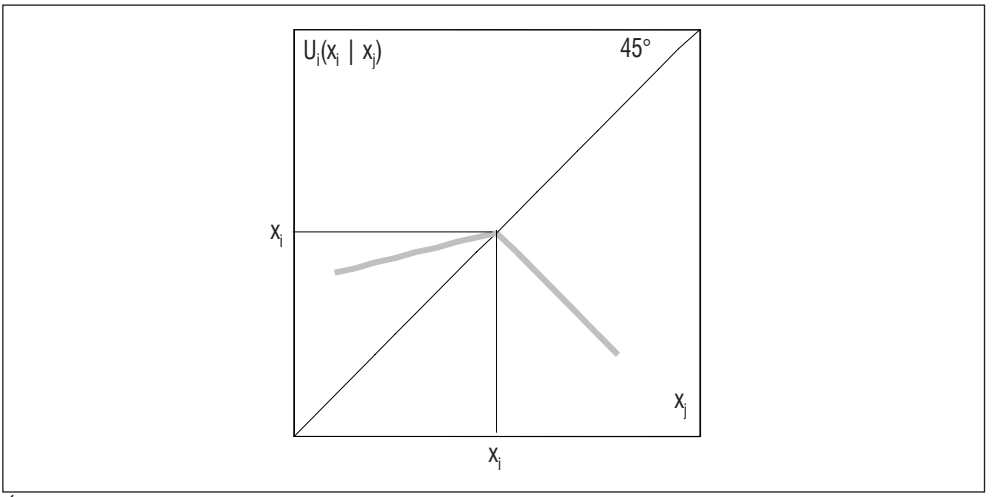
gdzie

$N = \{1,...n\}$ – zbiór graczy,
x – wektor końcowych materialnych wypłat w grze,
x_i – wypłata gracza *i* dla $i \in \{1,...,n\}$,
$\beta_i \in \langle 0;1)$,
$\alpha_i \in \langle \beta_i; +\infty)$.

E. Fehr i K. Schmidt przyjmują, że na użyteczność gracza *i* większy wpływ ma sytuacja, w której to on dostaje niższe wypłaty niż inni, niż sytuacja odwrotna. Stąd: $\beta_i \leq \alpha_i$. Oznacza to, że w dziedzinie społecznych porównań gracze wykazują *awersję do straty* – negatywne odchylenia otrzymywanej przez nich wypłaty od wypłat innych graczy w większym stopniu wpływają na ich użyteczność niż odchylenia pozytywne. W modelu zakłada się także, że $0 \leq \beta_i < 1$, tzn. decydenci nie czerpią przyjemności z tego, że są lepiej materialnie sytuowani niż inni. Przyjęcie, że $\beta_i < 1$ wyklucza dodatkowo sytuację, w której gracz *i* byłby gotowy zniszczyć część przypadających mu zasobów, by zredukować swoją przewagę nad innymi graczami. Awersja do nierówności odczuwana przez gracza *i* jest niezależna od liczby graczy oraz jest *egocentryczna*, tzn. gracz *i* dba przede wszystkim o relacje między wielkością własnej materialnej wypłaty a wypłatami każdego z pozostałych $n - 1$ graczy, nie interesują go natomiast nierówności między materialnymi wypłatami pozostałych uczestników gry. Rysunek 1 przedstawia użyteczność gracza *i* w grze dwuosobowej jako funkcję materialnej wypłaty otrzymywanej przez gracza *j*. Dla danego x_i użyteczność gracza *i* osiąga maksimum, gdy gracz *j* otrzymuje wypłatę równą wypłacie gracza *i* ($x_i = x_j$), użyteczność gracza *i* maleje, gdy jego wypłata jest wyższa niż wypłata gracza *j* (ma on z tego powodu poczucie winy) oraz gdy jego wypłata jest niż niższa niż wypłata gracza *j* (w tej sytuacji odczuwa on zazdrość).

G. Bolton i A. Ockenfels (2000) w swoim modelu za czynnik motywujący wybory danego gracza oprócz wypłaty materialnej x_i uznają także relatywną wypłatę otrzymywaną przez gracza *i*. W modelu zakłada się, że gracze porównują otrzymaną przez siebie wypłatę z całkowitą wypłatą otrzymaną przez innych graczy, a nie z wypłatami otrzymywanymi przez każdego z pozostałych graczy z osobna. Autorzy podkreślają, że model ten pozwala na racjonalizację trzech typów zachowań obserwowanych w różnych eksperymentach ekonomicznych: zachowań motywowanych troską o sprawie-

Rysunek 1. Użyteczność gracza *i* w modelu Fehra i Schmidta jako funkcja wypłaty pieniężnej dla gracza *j*



Źródło: Fehr, Schmidt (1999: 823).

dliwość podziału (ang. ***e**quity*), zachowań motywowanych chęcią odwzajemniania działań innych graczy (ang. ***r**eciprocity*) oraz zachowań ściśle konkurencyjnych (ang. ***c**ompetition*). Z tego względu nadają swojemu modelowi nazwę ERC.

Funkcja użyteczności (u_i) gracza *i* w modelu G. Boltona i A. Ockenfelsa ma następująca postać:

$$u_i = u_i(x_i, \sigma_i) \qquad [2]$$

przy czym:

$$\sigma_i(x_i, c, n) = \begin{cases} x_i / c & \text{dla } c > 0 \\ 1/n & \text{dla } c = 0 \end{cases}, \qquad [3]$$

gdzie:

$$c = \sum_{j=1}^{n} x_i \ . \qquad [4]$$

Zgodnie z przyjmowanymi przez autorów założeniami (Por. Bolton, Ockenfels 2000: 171-2) dla danego σ gracz *i* dąży do maksymalizacji swojej wypłaty pieniężnej w grze. Funkcja u_i nie musi być jednak ściśle rosnąca ze względu na wypłatę gracza *i*. Autorzy przyjmują bowiem, że $u_i^1(x_i;\sigma_i) \geq 0$[7]. W ten sposób uwzględnia się możliwość, że nie-

którzy gracze większą wagę przykładać będą do wypłaty relatywnej (σ_i) niż do wypłaty bezpośredniej x_i. Autorzy zakładają także, że $u_i^2(x_i;\sigma_i) = 0$ dla $\sigma_i(x_i,c,n) = 1/n$ oraz $u_i^{22} < 0$. Dla danej wielkości x_i funkcja u_i osiąga więc maksimum, gdy udział materialnej wypłaty danego gracza w sumie wypłat z gry jest równy $1/n$, czyli gdy wypłata gracza *i* jest równa średniej wypłacie w grze. Inaczej niż w modelu Fehra i Schmidta, w którym dany gracz, oceniając wynik gry porównuje własną materialną wypłatę bezpośrednio z wypłatami pozostałych graczy, w modelu ERC indywidualna wypłata porównywana jest tylko z wypłatą średnią. Jeżeli więc wypłata pieniężna gracza *i* będzie poniżej wartości średniej dla całej grupy, to będzie on chciał zmniejszyć wypłatę gracza *j* nawet wówczas, gdy wypłata gracza *j* będzie znacznie niższa od jego własnej wypłaty. Ponadto przy danej wielkości x_i gracz *i* będzie miał taką samą użyteczność w sytuacji, gdy wypłaty pieniężne w grze podzielone będą równo między wszystkich graczy, jak i wówczas, gdy rozkład wypłat będzie nierówny, ale gracz ten otrzyma wypłatę równą średniej wypłacie w grze.

Zaproponowany przez G. Charnessa i M. Rabina (2002) model preferencji quasi-maksiminowych stanowi syntezę podejścia opartego na awersji do niesprawiedliwych wyników gier z troską o efektywność podziału. Użyteczność gracza *i* w modelu G. Charnessa i M. Rabina jest średnią ważoną komponentu, który przez autorów nazywany jest *„bezstronnym kryterium dobrobytu społecznego"* (ang. *disinterested social welfare criterion*) oraz indywidualnej wypłaty pieniężnej gracza. *„Bezinteresowne kryterium dobrobytu społecznego"* to natomiast średnia ważona preferencji Utylitarnych oraz preferencji Rawlsowskich. Preferencje Utylitarne wyróżnione są ze względu na *zasadę utylitarystów*, nakazującą maksymalizację sumy wypłat pieniężnych w grze, preferencje Rawlsowskie – w oparciu o *kryterium maksyminu Rawlsa*, które nakazuje wybór tej spośród możliwych końcowych alokacji wypłat pieniężnych w grze, która maksymalizuje wypłatę osoby w najmniej korzystnej sytuacji. Uzależnienie użyteczności graczy od wartości *„bezinteresownego kryterium dobrobytu społecznego"* stanowi wyraz idei, zgodnie z którą *„gracze chcą pomagać wszystkim, ale są szczególnie skłonni do tego, by pomagać osobom w najmniej korzystnym położeniu"* (Charness, Rabin 2002: 821). Formalnie, funkcja użyteczności gracza *i* ($v_i(\mathbf{x})$) zgodnie z modelem preferencji quasi-maksiminowych ma następującą postać:

$$v_i(\mathbf{x}) = (1-\lambda)x_i + \lambda[\delta \min[x_1,...,x_N] + (1-\delta)[x_1 + ... + x_N]] \qquad [5]$$

gdzie:

x – wektor końcowych materialnych wypłat graczy,

x_i – wypłata końcowe gracza *i*, $i \in \{1,...,N\}$,

[7] Dalej przyjmuje się następujące oznaczenia: u_i^j – pochodna pierwszego rzędu ze względu na argument *j* funkcji użyteczności, u_i^{jj} – pochodną drugiego rzędu ze względu na argument *j* funkcji.

$\delta \in (0,1)$,
$\lambda \in \langle 0,1 \rangle$.

Parametr λ mierzy, w jakim stopniu na użyteczność gracza wpływa dobrobyt innych osób, a w jakim jego własna wypłata. Parametr δ mierzy natomiast, w jakim stopniu na wartość *„bezinteresownego kryterium dobrobytu społecznego"* wpływa użyteczność osoby w najgorszym położeniu, a w jakim całkowita suma wypłat pieniężnych w grze. Łatwo można zauważyć, że w grach dwuosobowych krańcowa użyteczność gracza *i* z wypłaty gracza *j* jest zawsze dodatnia. Zgodnie z założeniami modelu G. Charnessa i M. Rabina gracz *i* nigdy nie czerpie więc satysfakcji ze zmniejszenia wypłaty gracza *j*. Ponadto krańcowa użyteczność gracza *i* z wypłaty gracza *j* jest mniejsza, gdy $x_i < x_j$ niż w przeciwnym przypadku. Gracz *i* dba o dobrobyt gracza *j* w mniejszym stopniu, gdy gracz *j* jest lepiej sytuowany niż on sam.

W modelu Charnessa i Rabina przyjmuje się, że wartości parametrów λ oraz δ są takie same dla wszystkich graczy. Tym samym, inaczej niż w modelach Fehra i Schmidta oraz Boltona i Ockenfelsa, nie dopuszcza się tu możliwości, by gracze mieli heterogeniczne preferencje.

3. Testowanie dystrybucyjnych modeli nieegoistycznych preferencji

W ostatnich latach przeprowadzonych zostało wiele eksperymentów, których celem była relatywna ocena konkurencyjnych teorii nieegoistycznych preferencji. Eksperymenty te projektowane są w taki sposób, by możliwa była odpowiedź na wiele szczegółowych pytań, wskazujących, który z modeli preferencji lepiej radzi sobie z wyjaśnianiem zachowań graczy obserwowanych w różnych typach gier.

3.1. Awersja do nierównych wyników podziału a efektywność podziału

Jeden z istotnych kierunków badań eksperymentalnych testujących dystrybucyjne modele nieegoistycznych preferencji próbuje udzielić odpowiedzi na pytanie o relatywne znaczenie dwóch czynników – awersji do nierównych wyników podziału i troski o efektywność podziału[8]. Zgodnie z modelami awersji do nierównych wyników podziału bardzo nierówne podziały będą odrzucane także wówczas, gdy takie działanie zmniejszać będzie całkowitą sumę materialnych wypłat w grze. Nato-

[8] W badaniach tych, a także w niniejszej pracy, podział najbardziej efektywny to taki, który maksymalizuje sumę wypłat materialnych w grze.

miast w myśl modeli, takich jak np. model preferencji quasi-maksiminowych, gracze mogą tolerować pewien wzrost nierówności, jeżeli prowadzi on do wzrostu sumy wypłat otrzymywanych przez graczy.

D. Engelmann i M. Strobel (2004) zaprojektowali serię eksperymentów, w których testowano implikacje tych dwóch konkurencyjnych założeń. Uczestnicy eksperymentów dokonywali wyboru spośród trzech alokacji, które przyznawały określoną liczbę punktów eksperymentalnych trzem osobom. Następnie uczestników przydzielono losowo do trzyosobowych grup oraz w losowy sposób przypisano im role Osoby 1, Osoby 2 lub Osoby 3. Przy obliczaniu wypłat pieniężnych dla poszczególnych uczestników eksperymentu pod uwagę brany był wyłącznie wybór Osoby 2. Uczestnicy zostali poinformowani o tej procedurze eksperymentalnej zanim dokonali wyborów spośród przedstawianych im alokacji. Wszystkie trzy każdorazowo rozpatrywane alokacje przyznawały Osobie 2 taką samą liczbę punktów. Wybory, jakich dokonali uczestnicy wylosowani do roli Osoby 2, wpływały więc tylko na materialne wypłaty pozostałych dwóch osób, nie miały natomiast wpływu na wielkość ich własnych wypłat. Przykładowe trzy warianty problemów decyzyjnych rozpatrywanych przez autorów przedstawiono w *tabeli 1*. W *wariancie I* model Fehra i Schmidta wskazuje na wybór alokacji A, która minimalizuje różnice między wypłatą Osoby 2 i wypłatami pozostałych osób. Alokacja ta jest jednocześnie alokacją, w której suma wypłat dla poszczególnych osób jest najwyższa. Jej wybór jest więc jednocześnie zgodny z postulatem efektywności podziału. Model ERC wskazuje natomiast na wybór alokacji C minimalizującej różnice między wypłatą Osoby 2 a wypłatą średnią. W eksperymencie przeprowadzonym przez Engelmanna i Strobla alokacja A wybierana była w ok. 87% przypadków, natomiast alokacja C tylko w ok. 7%. W *wariancie II* zarówno postulat efektywności podziału, jak i model ERC nakazują wybór alokacji A, natomiast model Fehra i Schmidta – alokacji C. W tym przypadku alokacja A była wybierana przez 40% wszystkich uczestników badania, zaś alokacja C przez ok. 43%. W *wariancie III* oba modele awersji do nierównych wyników podziału wskazują na wybór alokacji C, która jednocześnie minimalizuje różnice między wypłatą decydenta a wypłatami pozostałych dwóch graczy oraz między wypłatą decydenta a średnią wypłatą w grze. Maksymalizacja sumy wypłat pieniężnych wskazywałaby natomiast na wybór alokacji A. Wyboru zgodnego z modelami awersji do nierówności dokonało w tym przypadku ok. 33% osób a z postulatem efektywności ok. 60%.

Analizując przytoczone powyżej wyniki, a także wyniki pozostałych wariantów problemów decyzyjnych rozpatrywanych w eksperymencie, D. Engelmann i M. Stro-

[9] W standardowej wersji gry *Dyktator* gracz A decyduje o podziale pewnej sumy pieniędzy między siebie i gracza B. Inaczej niż w przypadku gry *Ultimatum*, gracz B nie ma wpływu na końcowy wynik – propozycja gracza A jest automatycznie wprowadzana w życie.

Tabela 1. Przykładowe warianty problemów decyzyjnych rozpatrywanych w eksperymencie D. Engelmanna i M. Strobla

	Wariant I			*Wariant II*			*Wariant III*		
	Alokacja A	*Alokacja B*	*Alokacja C*	*Alokacja A*	*Alokacja B*	*Alokacja C*	*Alokacja A*	*Alokacja B*	*Alokacja C*
Wypłata Osoby 1	17	18	19	21	17	13	14	11	8
Wypłata Osoby 2 (decydenta)	**10**	**10**	**10**	**12**	**12**	**12**	**4**	**4**	**4**
Wypłata Osoby 3	9	5	1	3	4	5	5	6	7
Średnia wypłat osób 1 i 3	13	11,5	10	12	10,5	9	9,5	8,5	7,5
Całkowita wypłata w grze	36	33	30	36	33	30	23	21	19

Źródło: Engelmann, Strobel (2004).

bel zauważyli, że odsetek zachowań wyjaśnianych przez modele awersji do nierówności jest znacznie niższy w sytuacji, gdy wybory implikowane przez te modele stoją w konflikcie z troską o efektywność podziału, niż gdy konflikt taki nie występuje.

Relatywne znaczenie takich czynników, jak troska o efektywność podziału oraz troska o jego równość, badali także G. Charness i M. Rabin (2002). Autorzy ci rozpatrywali grę *Dyktat*[9], w której osoba decydująca o podziale dokonuje wyboru spośród dwóch dostępnych alokacji. Alokacja I przyznaje obu osobom po 400 jednostek. W alokacji II decydent otrzymuje 400, a druga osoba 750 jednostek. Okazało się, że ponad 2/3 decydentów wybierało alokację II. Alokacja ta zwiększa nierówności między uczestnikami, zwiększając zarazem całkowitą materialną wypłatę w grze. Ponieważ w grupie osób wybierających alokację II mogły być zarówno osoby maksymalizujące sumę wypłat w grze, jak i osoby z preferencjami egoistycznymi, przeprowadzony został dodatkowy eksperyment. W eksperymencie tym decydenci wybierali między alokacją III przyznającą obu osobom 400 jednostek oraz alokacją IV, która decydentowi przyznawała 375 jednostek, natomiast drugiemu uczestnikowi 725 jednostek. W tym wariancie podział maksymalizujący dzieloną kwotę wybierany był przez ok. połowę „dyktatorów". Wyniki te świadczą o tym, że troska o efektywność może w istotnym stopniu wpływać na zachowania decydentów. Wyniki ankiety przeprowadzonej dodatkowo przez autorów wśród uczestników eksperymentów sugerują jednak, że znaczenie tego czynnika może maleć wraz ze wzrostem nierówności między graczami.

Dowodów na to, że w obliczu konfliktu między równością a efektywnością podziału przeważać może drugi czynnik, dostarczają także eksperymenty, jakie przeprowadzili J. Cox i V. Sadiraj (2006). Eksperymenty te oparte były na pewnej modyfikacji gry *Dyktat*. W grach tych zarówno osoba wchodząca w rolę „dyktatora", jak i osoba, z którą „dyktator" może podzielić się posiadanymi zasobami, początkowo posiadają po 10 jednostek waluty eksperymentalnej. Kwota przekazywana przez „dyktatora" drugiej osobie mnożona jest przez współczynnik $m = 3$. Teorie preferencji opartych

Tabela 2. Wypłaty materialne w czteroosobowej grze Dyktat w eksperymencie J. Coxa i V. Sadiraj

	Alokacja A	***Alokacja B***	***Alokacja C***
Wypłata gracza 1	5	5	5
Wypłata gracza 2 (decydenta)	**15**	**15**	**15**
Wypłata gracza 3	11	20	7
Wypłata gracza 4	11	20	38

Źródło: Cox, Sadiraj (2006).

na awersji do nierównych wyników podziału przewidują w tym wypadku brak jakichkolwiek transferów, gdyż każdy transfer zwiększał będzie nierówność. Jednak aż 63% uczestników, którzy odgrywali rolę „dyktatora", przekazywało część posiadanych przez siebie zasobów. Ci sami autorzy pokazali jednak, że troska o efektywność w niektórych sytuacjach może nie być w stanie tłumaczyć wyborów dokonywanych przez uczestników eksperymentów. Działo się tak między innymi w przypadku eksperymentu opartego na czteroosobowej grze *Dyktat*, w której „dyktator" decydował o wyborze jednej z trzech dostępnych alokacji wypłat materialnych między czterech uczestników eksperymentu. We wszystkich trzech alokacjach wypłata dla decydenta i wypłata minimalna były takie same (*tabela 2*). Troska o efektywność podziału nakazywałaby więc wybór alokacji C, która maksymalizowała całkowitą sumę materialnych wypłat uczestników gry. Alokacja ta wybierana była jednak tylko przez ok. 15% „dyktatorów", aż 85% „dyktatorów" wybierało „bardziej równe" alokacje A oraz B.

Przedstawione w artykule eksperymenty testujące relatywne znaczenie efektywności i awersji do nierówności pokazały, że efektywność może odgrywać ważną rolę w motywowaniu wyborów dokonywanych przez część uczestników tych eksperymentów. Niektórzy badani skłonni są poświęcić część swoich zasobów w celu zwiększenia sumy wypłat pieniężnych nawet wówczas, gdy konsekwencją tego jest wzrost nierówności między uczestnikami podziału. Inni nad „efektywność podziału" przedkładają jego równość. Wyniki tych eksperymentów nie pozwalają więc jednoznacznie rozstrzygnąć o tym, które modele – te oparte na awersji do nierówności, czy te uwzględniające efektywność podziału – lepiej opisują zachowania uczestników badań. Sugerują raczej, że żaden z omawianych modeli nie jest w stanie racjonalizować wszystkich obserwowanych w eksperymentach zachowań, wskazując tym samym na heterogeniczność preferencji badanych osób. Dodatkowo należy podkreślić, że przeważająca większość eksperymentów testujących relatywne znaczenie efektywności i awersji do nierówności (w tym wszystkie omawiane w tym artykule) oparta jest na grze *Dyktat*, z czym wiążą się dwa istotne problemy. Po pierwsze, wyniki uzyskiwane w eksperymentach opartych na grze *Dyktat* są bardzo wrażliwe na wszelkie zmiany w warunkach eksperymentalnych (Fehr, Schmidt 2006). Po drugie, gra *Dyktat* zakłada brak strategicznych interakcji między uczestnikami eksperymentu. Oso-

by niebędące „dyktatorami" nie mają możliwości wpływu na wypłaty osoby decydującej o wyniku gry. W związku z tym pojawia się pytanie, czy troska o efektywność podziału nie jest wyłącznie rezultatem eliminacji pewnego istotnego aspektu strategicznego, jakim jest interakcja między graczami, oraz czy w warunkach, w których znaczenie mają zachowania strategiczne, postrzegane intencje czy uczenie się, inne motywacje nie okażą się bardziej istotne (Fehr, Schmidt 2006).

3.2. Testowanie modeli awersji do nierównych wyników podziału

Znaczna liczba badań eksperymentalnych poświęcona jest próbie rozstrzygnięcia między dwoma modelami opartymi na awersji do nierównych wyników podziału: modelem Fehra i Schmidta oraz modelem Boltona i Ockenfelsa. Modele te różnią się między sobą przede wszystkim założeniem dotyczącym materialnych wypłat, z jakimi dany gracz porównuje własną wypłatę. Przypomnijmy, że w modelu Fehra i Schmidta są to pieniężne wypłaty wszystkich graczy z osobna, w modelu Boltona i Ockenfelsa – średnia wypłata w grze.

G. Charness i M. Rabin (2002) przetestowali te konkurencyjne założenia obu modeli za pomocą bardzo prostego eksperymentu. W eksperymencie tym gracz oznaczony jako „C" mógł dokonać wyboru spośród dwóch możliwych alokacji. Alokacja I przyznawała wszystkim trzem graczom po 575. Alokacja II przyznawała graczowi A – 900, B – 300 oraz C – 600. Zauważmy, że w obu przypadkach gracz C otrzymywał wypłatę równą średniej wypłacie w grze. Zgodnie z modelem ERC powinien więc wybierać alokację II, jako że w ujęciu absolutnym dawała mu ona wyższą materialną wypłatę niż alokacja I. Uczestnik, który dokonuje porównań własnej wypłaty z wypłatami innych w sposób opisany przez model Fehra i Schmidta, powinien natomiast wybierać alokację I. W badaniu G. Charnessa i M. Rabina pierwszą alokację wybierało 54% osób, drugą – 46%. Wynik ten nie pozwala jednoznacznie rozstrzygnąć, które z założeń dotyczących sposobu porównywania wypłat przez graczy, lepiej opisuje faktyczne zachowania uczestników eksperymentu. Warto jednak zauważyć, że spośród 46%. osób, które wybrały alokację II część osób mogła mieć preferencje *stricte* egoistyczne, a ten typ preferencji jest zgodny z oboma modelami awersji do nierównych wyników podziału.

Inny prosty eksperyment pozwalający na testowanie założeń modeli awersji do nierównych wyników podziału zaproponowali E. Fehr i U. Fischbacher (2004). W grach eksperymentalnych każdorazowo uczestniczyło trzech graczy. Między graczami A oraz B rozgrywana była gra *Dyktat*, przy czym o podziale 100 jednostek waluty eksperymentalnej decydował gracz A. Gracz C obserwował decyzję gracza A i mógł wykorzystać 50 jednostek, jakimi dysponował, do ukarania gracza A. Każda jednostka wydana przez gracza C redukowała wypłatę materialną gracza A o 3 jednostki. Gracz A wiedział, że

może zostać ukarany przez gracza C. Gdy gracz C nie decydował się na ukaranie gracza A, jego wypłata wynosiła 50 jednostek, co stanowiło dokładnie 1/3 kwoty, która dzielona była między uczestników badania. Zgodnie z modelem ERC gracz C nie powinien więc karać gracza A nawet wówczas, gdy ten proponował bardzo nierówny podział. Badanie pokazało jednak, że aż 60% graczy typu C wymierza karę graczom A, jeżeli ci nie są skłonni dzielić się z graczami typu B. Ponadto, im mniej przekazywali gracze A, tym większa była przeciętna wielkość kary wymierzanej przez gracza C. Zachowania takie zgodne są z modelem Fehra i Schmidta.

Wyniki przywołanych wyżej eksperymentów sugerują, że oceniając różne alokacje, uczestnicy eksperymentów biorą pod uwagę pieniężne wypłaty pozostałych graczy z osobna, a nie średnią wypłatę w grze. Świadczy to na korzyść modelu zaproponowanego przez E. Fehra i K. Schmidta i na niekorzyść modelu ERC. Warto jednak zauważyć, że większość gier eksperymentalnych rozgrywana jest w małych grupach, w których porównywanie własnej materialnej wypłaty z wypłatami wszystkich pozostałych graczy nie nastręcza graczom istotnych trudności. Być może w dużych grupach, w których porównania takie wymagałyby znacznego wysiłku, model ERC byłby bardziej adekwatny do opisu zachowań graczy.

4. Krytyka dystrybucyjnych modeli nieegoistycznych preferencji

Wobec dystrybucyjnych modeli nieegoistycznych preferencji wysuwanych jest wiele zastrzeżeń (Sobel 2005: 430-432). Po pierwsze, teoriom tym, a także wszelkim innym teoriom, które rozszerzają funkcje użyteczności o jakieś dodatkowe zmienne, zarzuca się, że poprzez takie modyfikacje funkcji użyteczności można w zasadzie „wyjaśnić wszystko, a więc nic". Każde zachowanie daje się bowiem wyjaśnić poprzez przyjęcie założenia, że jest to zachowanie preferowane. W przypadku teorii nieegoistycznych preferencji zarzut o „niekontrolowanych manipulacjach" funkcjami użyteczności nie jest jednak zasadny. Teorie te powstawały w reakcji na wyniki eksperymentów dociekających, jakie czynniki motywują konkretne wybory i zachowania uczestników tych eksperymentów. Wprowadzanie nowych komponentów do funkcji użyteczności w przypadku teorii nieegoistycznych preferencji nie odbywało się więc w sposób arbitralny, lecz miało swoje oparcie w empirii.

[10] Badanie tego typu przeprowadzili także J. Brosig, T. Reichmann, J. Weimann (2007).

[11] W tym miejscu chciałabym szczególnie podziękować anonimowemu recenzentowi za jego komentarz zgłoszony w odniesieniu do tekstu Blanco, Engelmann, Norman (2007), a także za wiele innych uwag i komentarzy zgłoszonych do niniejszego artykułu. W oparciu o ten komentarz zredagowany został fragment artykułu poświęcony pracy Blanco et. al.

Drugi zarzut dotyczy zbyt mało restrykcyjnych ograniczeń nakładanych na parametry dystrybucyjnych modeli nieegoistycznych preferencji. Dobre dopasowanie modeli nieegoistycznych preferencji do danych eksperymentalnych często przypisywane jest temu, że modele te mają dużą liczbę parametrów, którymi można swobodnie manipulować w celu uzyskania zgodności między przewidywaniami modelu a wynikami eksperymentów. Zarzut ten wysuwany jest przede wszystkim pod adresem modeli preferencji dopuszczających heterogeniczne preferencje graczy (czyli np. wobec modelu Fehra i Schmidta oraz modelu ERC). W przypadku tych teorii końcowy wynik gry często zależy bowiem od rozkładu osób z różnymi typami preferencji w populacji graczy. Dla każdego rezultatu otrzymanego w badaniu eksperymentalnym można więc dobrać taki rozkład parametrów indywidualnych funkcji użyteczności, by rezultat ten dał się objaśniać przy pomocy danego modelu. Pojawia się więc problem z falsyfikowalnością tych teorii. W tym kontekście istotne jest jednak to, by odróżniać predykcje teorii od predykcji formułowanych na podstawie konkretnych parametryzacji modeli nieegoistycznych preferencji. W przypadku tych drugich możliwe jest bowiem prowadzenie bardziej rygorystycznych testów. Badania takie w odniesieniu do modelu Fehra i Schmidta przeprowadzili przykładowo M. Blanco, D. Engelmann, H. Norman (2007)[10,11]. Autorzy ci zauważyli, że o ile na poziomie zagregowanym poddana testowaniu parametryzacja modelu Fehra i Schmidta potrafi dobrze wyjaśniać prawidłowości obserwowane w różnych typach gier (m.in. w grach: *Ultimatum*, *Dylemat Więźnia*, *Dyktat*), o tyle na poziomie jednostkowym parametryzacja modelu oszacowana na podstawie wyników jednej gry nie pozwala na przewidywanie zachowań uczestnika eksperymentu w innych rozpatrywanych grach. Zdaniem autorów zaobserwowany przez nich brak statystycznej zależności między przypisanymi graczom parametryzacjom modelu preferencji, a ich zachowaniem w grach, w których testowane były przewidywania formułowane w oparciu o te parametryzacje, wskazuje na „niską korelację motywacji między grami". Oznacza to, że motywacje stojące za wyborami dokonywanymi przez graczy mogą różnić się w zależności od kontekstu strategicznego tych wyborów. Ta zależność motywacji od kontekstu strategicznego stawia pod znakiem zapytania możliwość opisu preferencji w różnych grach za pomocą dostępnych obecnie modeli nieegoistycznych preferencji. Zaobserwowana przez Blanco et. al. niezgodność między przewidywaniami jednego z tych modeli a danymi empirycznymi nie musi oznaczać jednak definitywnej porażki modeli nieegoistycznych preferencji. Badania nad modelami nieegoistycznych preferencji stanowią stosunkowo nową dziedzinę badań eksperymentalnych, zaś same modele nie stanowią jeszcze zamkniętych rozstrzygnięć, lecz pewne wstępne propozycje, które mogą być modyfikowane w oparciu o wyniki testujących je eksperymentów. Rozbieżności między przewidywaniami modeli nieegoistycznych preferencji a zachowaniem uczestników eksperymentów, na które wskazuje m.in. praca Blanco et. al., będą prawdopodobnie wyznaczać nowe kierunki badań nad modelami nieegoistycznych preferencji.

Kolejny zarzut dotyczy modeli, które jako kryterium oceny sprawiedliwości podziału przyjmują zasadę równości podziału, a dokładniej równości końcowych materialnych wypłat graczy. Równy podział jest tylko jedną z wielu zasad sprawiedliwości dystrybutywnej, często nie najważniejszą. Zasada ta może jednak dominować w sytuacjach ubogich w kontekst, w których nie ma dużo informacji na temat sytuacji materialnej graczy, ich potrzeb czy też uprawnień. Jednakże zasada ta może zostać zdominowana przez inne zasady wówczas, gdy informacje takie są powszechnie dostępne. Tym samym modele oparte na tej zasadzie, mimo przydatności do wyjaśniania zachowań uczestników eksperymentów ekonomicznych, mogą okazać się zbyt uproszczonym narzędziem do analizy danych generowanych w bardziej rzeczywistych kontekstach. Istotnie, badania terenowe, a także niektóre badania eksperymentalne wskazują, że preferencje dystrybucyjne często oparte są na innych zasadach sprawiedliwego podziału niż zasada równości wypłat pieniężnych (Konow 2000).

5. Znaczenie intencji graczy

Modele dystrybucyjne stanowiące główny przedmiot niniejszego artykułu zakładają, że wybory uczestników eksperymentów dokonywane są na podstawie dokonywanej przez nich oceny końcowego rozkładu wypłat materialnych w rozgrywanej grze. Autorzy tych modeli zakładają, że dzieje się tak zarówno w kontekstach niestrategicznych (np. dających się opisać za pomocą gry *Dyktat*), jak i w kontekstach strategicznych[12]. Jednakże można argumentować, że *równość* wypłat pieniężnych w grze nie jest dobrym kryterium oceny sprawiedliwości tej gry; że zachowania graczy w kontekstach strategicznych zależą nie tylko od tego, jaka będzie końcowa alokacja pieniędzy w grze, ale także na przykład od tego, jak oceniają oni zachowania i intencje innych graczy. Taka intuicja leży u podstaw *modeli wzajemności opartej na ocenie intencji graczy*[13]. W modelach tych wybór akcji przez gracza uzależniony jest od tego, jak ocenia on zachowania i intencje innych graczy. Przyjmuje się tu, że gracz chce odpowiadać przyjaźnie na te działania innych osób, które postrzega jako przyjazne oraz że chce karać te zachowania, które ocenia jako wrogie (Rabin 1993: 1282). Z kolei to, czy dane działanie

[12] Zarówno E. Fehr i K. Schmidt (1999), jak i G. Bolton i A. Oeckenfels (2000) pokazują, że przy pomocy ich modeli można wyjaśniać wiele prawidłowości obserwowanych w kontekstach strategicznych, które dają się opisać za pomocą takich gier, jak np. gra *Ultimatum*, *Dylemat Więźnia* czy *Produkcja Dóbr Publicznych*.

[13] W klasie modeli wzajemności wyróżnić można: ***1) modele wzajemności opartej na typie***, w których to, jakie działania będzie podejmował dany gracz w odpowiedzi na działania innych graczy zależało będzie od oceny „charakteru" (typu) tych graczy (np. Levine 1998); ***2) modele wzajemności opartej na ocenie intencji*** (Rabin 1993; Dufwenberg, Kirchsteiger 2004; Falk, Fischbacher 2006); ***3) modele wzajemności opartej na porównaniu faktycznych zachowań graczy z pewnym standardem porównań*** (Erlei 2004; Charness; Rabin 2002; Cox i inni 2007).

postrzegane jest jako przyjazne, czy też nie, zależy od oceny intencji graczy, którzy działanie to podejmują. Chcąc oceniać intencje, musimy posiadać pewne przekonania zarówno na temat tego, co dany gracz planuje zrobić, jak i na temat tego, dlaczego planuje on to zrobić. Aby udzielić odpowiedzi na oba te pytania, musimy mieć pewne przekonania na temat tego, co ów gracz zakłada odnośnie naszych planów. Inaczej będziemy bowiem oceniać jego intencje wówczas, gdy jesteśmy przekonani, że on wierzy, że zachowamy się wobec niego przyjaźnie. Inaczej zaś wówczas, gdy wierzymy, że jest on przekonany, że będziemy w stosunku do niego złośliwi. Oceniając intencje, porównujemy wypłaty pieniężne, do których prowadzi dana akcja gracza, z którym rozgrywamy pewną grę, z wypłatami, do których mogłyby doprowadzić inne akcje, uwzględniając jednocześnie to, że osoba ta może mieć pewne oczekiwania dotyczące akcji, którą my wybierzemy. Do modelowania tego typu zagadnień autorzy modeli wzajemności opartej na ocenie intencji wykorzystują narzędzia *psychologicznej teorii gier*, rozwijanej przez J. Geanakoplosa, D. Pearce'a i E. Stacchettiego (1989). Pierwszy model wzajemności opartej na ocenie intencji zaproponowany został przez M. Rabina (1993). Model ten, sformułowany dla przypadku dwuosobowych gier w postaci normalnej ze skończonym zbiorem strategii, został następnie uogólniony dla przypadku gier *n*-osobowych w postaci sekwencyjnej przez M. Dufwenberga i G. Kirschteigera (2004) oraz przez A. Falka i U. Fischbachera (2006). W tym ostatnim modelu przyjęto dodatkowo, że dla oceny intencji innych graczy znaczenie ma to, czy alokacja wypłat pieniężnych między poszczególnych uczestników gry jest sprawiedliwa. Przy czym, jako kryterium oceny sprawiedliwości przyjęta zostało kryterium *równości* materialnych wypłat w grze. Model A. Falka i U. Fischbachera stanowi więc syntezę modeli wzajemności opartej na ocenie intencji oraz modeli awersji do niesprawiedliwych wyników podziału.

Wprowadzanie do modeli nieegoistycznych preferencji *idei wzajemności* jest zgodne z bardzo prostymi intuicjami dotyczącymi motywów leżących u podstaw ludzkich zachowań. Jednakże konsekwencją uwzględnienia *wzajemności* jest znaczna złożoność tych modeli[14]. Modele te dopuszczają zazwyczaj istnienie wielu równowag, co utrudnia testowanie formułowanych na ich podstawie hipotez odnośnie zachowań uczestników gier eksperymentalnych. Ponadto analiza nawet prostych gier za pomocą tego typu modeli preferencji jest dość wymagającym zadaniem. Dlatego też modele te nie są tak często testowane w eksperymentach, jak modele dystrybucyjne. Nieliczne eksperymenty pozwalające na testowanie założeń modeli wzajemności opartej na ocenie intencji graczy usiłują między innymi rozstrzygnąć, co motywuje odrzucanie skrajnie nierównych podziałów przez badane osoby – czysta awersja do nierówne-

[14] Ze względu na tą złożoność modele te nie są omawiane w niniejszym artykule. Dobre przedstawienie tych modeli wymagałoby prawdopodobnie poświęcenia im odrębnego artykułu. Czytelnicy zainteresowani tematyką odsyłani są do artykułów źródłowych. Prezentację uproszczonych wersji tych modeli znaleźć można w pracy Camerera (2003).

go podziału czy też chęć zemsty na osobie, która dokonała niekorzystnego dla nich wyboru (*negatywna wzajemność*). Ocena intencji decydentów jest istotna tylko w tym drugim przypadku. Ten kierunek badań eksperymentalnych próbuje także określić, czy ocena intencji może wpływać na gotowość do podejmowania zachowania kooperatywnych i do poświęceń na rzecz innych (*pozytywna wzajemność*).

Kwestie te badał m.in. G. Charness (1996). Zaproponowany przez niego eksperyment przybliżał sytuację wymiany świadczeń między pracobiorcami a pracodawcą. Pracobiorcy otrzymywali wynagrodzenie, które w zależności od rozgrywanego wariantu ustalane było bądź w sposób losowy, bądź przez pracodawcę, bądź wreszcie przez jakiegoś zewnętrznego arbitra. Następnie wybierali oni poziom wysiłku, jaki chcieli wkładać w wykonywanie pracy na rzecz pracodawcy. Pracodawca na początku gry dysponował zasobem 120 punktów eksperymentalnych. Funkcje materialnych wypłat pracodawcy i pracobiorcy określone były w następujący sposób:

$$\Pi_F = (120 - w) \times e, \qquad [6]$$

$$\Pi_E = w - c(e) - 20, \qquad [7]$$

gdzie:

Π_F, Π_E – funkcje materialnych wypłat odpowiednio pracodawcy i pracobiorcy;

$w \in \langle 20; 120 \rangle$ – płaca, jaką pracodawca może zaoferować pracobiorcy;

$e = \{0{,}1; 0{,}2; 0{,}3; 0{,}4; 0{,}5; 0{,}6; 0{,}7; 0{,}8; 0{,}9; 1\}$ – poziom wysiłku, który może zostać wybrany przez pracownika;

$c(e)$ – funkcja kosztu będąca rosnącą funkcją poziomu wysiłku wybranego przez pracownika[15].

Załóżmy, że poszczególni gracze dążą do maksymalizacji swoich wypłat pieniężnych (tzn. mają egoistyczne preferencje). W takiej sytuacji pracownicy, dla których ponoszenie wysiłku jest kosztowne i którzy nie mogą być karani, jeżeli wybiorą niski poziom wysiłku, nie mają bodźców do tego, by wybierać poziom wysiłku przekraczający minimalny poziom 0,1. Jednakże w eksperymencie Charnessa tylko ok. 21% pracowników wybierało minimalny poziom wysiłku niezależnie od wielkości zaoferowanej im płacy. Pozostałe 79% pracowników było skłonnych zwiększać poziom wysiłku wraz ze wzrostem płacy. Zachowania takie są zgodne z przewidywaniami modeli awersji do nierównych wyników podziału, gdyż wybór minimalnego poziomu wysiłku prowadzi do coraz większych nierówności między wypłatą pracodawcy i pracobiorcy wraz ze wzrostem oferowanej płacy. Silny pozytywny zwią-

[15] W eksperymencie uczestnikom przedstawiono tabelę, w której poziomom wysiłku, które wybrać mogą pracobiorcy, przyporządkowano odpowiednią wielkość kosztu ponoszenia tego wysiłku. Wysiłkowi na poziomie 0,1 przyporządkowano koszt wynoszący 0 punktów eksperymentalnych, wysiłkowi na poziomie 1 – koszt rzędu 18 punktów eksperymentalnych.

zek między otrzymaną płacą a wybranym poziomem wysiłku zaobserwowano we wszystkich trzech wersjach eksperymentu, tzn. zarówno wówczas, gdy płace były ustalane przez pracodawcę, jak i wówczas, gdy były ustalane w sposób losowy bądź przez zewnętrznego arbitra. Znaczna część pracobiorców wykazywała więc gotowość do ponoszenia wysiłku na rzecz pracodawcy nawet wówczas, gdy to nie on odpowiadał za poziom ich płac, co jest zgodne z modelami opartymi na awersji do nierówności, ale nie z preferencjami opartymi na idei wzajemności. Pracownicy otrzymujący najniższe oferty płacowe wybierali jednak istotnie niższy poziom wysiłku wówczas, gdy poziom wynagrodzenia wyznaczał pracodawca. Wynik ten potwierdził więc tezę o tym, że gracze skłonni są karać zbyt egoistyczne zachowania innych osób (*negatywna wzajemność*). Nie udało się natomiast w sposób jednoznaczny potwierdzić tezy o istotnym znaczeniu *pozytywnej wzajemności*. W oparciu o rozkład ofert płacowych zaobserwowanych w eksperymencie Charness wyodrębnił te oferty, które mogły uchodzić za bardzo hojne. Następnie badał, czy poziom wysiłku wybieranego przez graczy otrzymujących takie oferty zależy od mechanizmu generowania decyzji dotyczących wysokości oferowanych płac (tzn. przykładowo od tego, czy ustalił je pracodawca czy też zostały ustalone w sposób losowy). W zależności od przyjętego progu, za pomocą którego wyodrębniano hojne oferty płacowe, różnice w poziomie wysiłku wybieranego przy różnych mechanizmach generowania decyzji okazywały się być albo istotne albo nieistotne statystycznie.

Ciekawy eksperyment testujący wpływ oceny intencji na decyzje podejmowane przez graczy zaprojektowali A. Falk, E. Fehr i U. Fischbacher (2008). Ich eksperyment bazuje na dwuosobowej grze *moonlighting*. Gra ta składa się z dwóch etapów. Na początku gry obaj gracze mają na koncie po 12 punktów. W pierwszym etapie gracz A wybiera wielkość $a \in \{-6,-5,...,5,6\}$, którą chce „przetransferować" do gracza B. Gdy $a > 0$, gracz A rezygnuje z a punktów na rzecz gracza B. Eksperymentator mnoży te punkty przez 3, tak że liczba dodatkowych punktów, które trafiają na konto gracza B, wynosi $3a$. Z konta gracza A odpisywane jest natomiast a punktów. Gdy gracz A wybiera $a < 0$, $|a|$ punktów odejmowane jest z konta gracza B i dopisywane na konto gracza A. W drugim etapie gry uprawnienia decyzyjne ma gracz B. Wyboru dokonać może spośród akcji $b \in \{-6,-5,...,17,18\}$. Dodatnia wartość b oznacza, że gracz B nagradza gracza A, przekazując mu b swoich punktów (b punktów transferowane jest z konta gracza B na konto gracza A). Ujemna wartość oznacza karę dla gracza A w wysokości $3|b|$ punktów, kara ta kosztuje gracza B dokładnie b punktów (z konta gracza A znika $3|b|$ punktów, z konta gracza B – b punktów). Autorzy badają zachowania graczy w dwóch wersjach eksperymentu. W pierwszej wersji (z intencjami) gracz A sam dokonuje wyboru wielkości a, w drugiej (bez intencji) wybór gracza A generowany jest w sposób losowy. Gracz B jest poinformowany odnośnie sposobu wyboru wielkości a.

Zgodnie z modelem egoistycznych preferencji gracz B nigdy nie będzie karał gracza A, wobec czego gracz A zawsze będzie wybierał $a = -6$. Modele awersji do nierównych wyników podziału wskazują, że b będzie rosnąć wraz ze wzrostem a oraz że $b = 0$ dla $a = 0$. Ponieważ w modelach tych intencje nie odgrywają żadnej roli, gracz B powinien zachowywać się tak samo w obu wersjach eksperymentu. Zgodnie z modelem M. Dufwenberga i G. Kirchsteigera, w wersji bez intencji gracz B nie powinien wykazywać ani negatywnego, ani pozytywnego odwzajemniania w stosunku do gracza A ($\forall a \quad b = 0$). W wersji eksperymentu uwzględniającej intencje model M. Dufwenberga i G. Kirchsteigera nie dostarcza jednak jednoznacznych przewidywań dotyczących zmian b w reakcji na zmiany a. Zgodnie z modelem Falka i Fischbachera, w przypadku eksperymentu z intencjami, b powinno być rosnącą funkcją a. Podobnej zależności, jednak o słabszym nasileniu, oczekuje się w wersji bez intencji. Model Falka i Fischbachera zakłada bowiem, że gracze reagują nie tylko na intencje innych, ale także na końcowy wynik podziału.

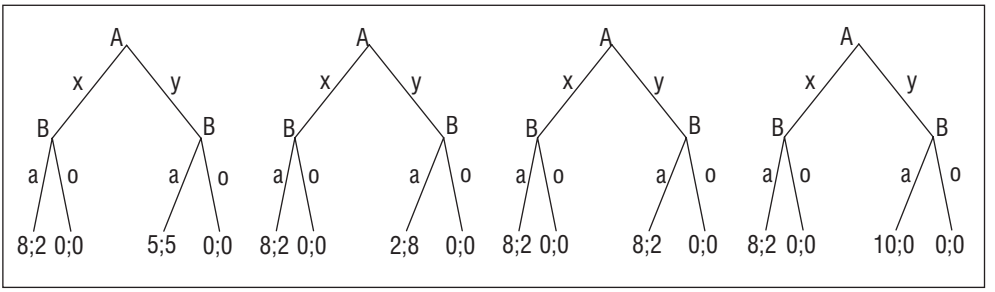
Wyniki badania eksperymentalnego pokazały, że ocena intencji innych graczy ma wpływ na wybory dokonywane przez uczestników eksperymentu. Zachowania obserwowane w poszczególnych wersjach eksperymentu istotnie się różniły, co jest sprzeczne z przewidywaniami modeli awersji do nierównych wyników podziału. W wersji z intencjami gracze B karali i nagradzali zachowania gracza A. Przeciętne wielkości nagród i kar rosły waz z wielkością a. W wersji bez intencji wielkość przeciętnych nagród i kar różniła się istotnie od zera tylko dla dostatecznie dużych wielkości a. Nagradzanie było „słabsze" niż w przypadku intencjonalnych działań gracza A, jednak nie zanikło zupełnie. Wydaje się więc, że pozytywne działania gracza B wobec gracza A wynikają nie tylko z oceny intencji gracza A, ale także z oceny końcowych rezultatów podziału. Wniosek taki zgodny jest z modelem Falka i Fischbachera, ale nie z modelem Dufwenberga i Kirchsteigera.

Podobne wnioski na temat poszczególnych modeli preferencji graczy formułują w oparciu o wyniki serii bardzo prostych eksperymentów związanych z grą *Ultimatum* A. Falk, E. Fehr, U. Fischbacher (2003). Gry rozważane przez autorów przedstawione zostały na *rysunku 2*. Każdorazowo gracz A wybiera między podziałem przyznającym mu 8 jednostek, a drugiemu graczowi 2 jednostki oraz jakimś innym podziałem. W zależności od rozpatrywanej gry są to podziały (10,0), (2,8), (5,5) oraz (8,2), gdzie pierwsza liczba oznacza wypłatę materialną dla gracza proponującego podział, druga – dla gracza, który propozycję tę przyjmuje bądź odrzuca. W ostatnim przypadku gracz A *de facto* nie ma wyboru i musi zaproponować graczowi B podział (8,2). Zauważmy, że modele preferencji nieegoistycznych oparte na końcowych rezultatach podziału przewidują identyczną, niezerową częstość odrzuceń propozycji podziału (8,2) we wszystkich czterech grach. Ponadto zauważmy, że „czystą" niechęć do nierównego podziału

mierzy gra, w której gracz A musi zaproponować podział (8,2). Taką awersję zaobserwowano u około 18% badanych. W pozostałych grach odsetek odrzuceń zaproponowanego podziału jest odpowiednio niższy lub wyższy, co jest sprzeczne z przewidywaniami modeli awersji do nierównych wyników podziału, w przypadku których niewybrane opcje z założenia nie mają wpływu na ocenę końcowych wyników gry. Proponowany podział jest znacznie częściej odrzucany w sytuacji, w której gracz A mógł wybrać równy podział 10 punktów. Odrzucenia występują rzadziej, gdy jedyną dostępną alternatywą był podział skrajnie nierówny. Rezultaty te zgodne są z przewidywaniami modelu A. Falka i U. Fischbachera, który bierze pod uwagę zarówno intencje, jak i końcowe wyniki podziału. Skłonność osób proponujących podział do wyboru podziału (8,2) także uzależniona jest od dostępnych alternatyw. Propozycje podzielenia się kwotą w proporcji 8:2 najrzadziej składane są w przypadku, gdy alternatywą jest podział równy, najczęściej – gdy alternatywą jest podział skrajnie nierówny.

Badania eksperymentalne wskazują więc, że modele oparte wyłącznie na końcowych materialnych wypłatach w grze (tj. omawiane w poprzednich rozdziałach modele dystrybucyjne) w wielu sytuacjach nie są w stanie wyjaśniać istotnej części zachowań graczy. Często znaczenie ma bowiem nie tylko ocena ostatecznych wypłat, ale i ocena intencji innych graczy. Wydaje się przy tym, że odwzajemnianie oparte na ocenie intencji innych graczy ma silniejszy wpływ na zachowania związane z karaniem nieżyczliwych niż nagradzaniem życzliwych decyzji.

Rysunek 2. Gry rozgrywane w eksperymencie A. Falka, E. Fehra, U. Fischbachera



Oznaczenia:
a – akceptacja podziału przez gracza B;
o – odrzucenie podziału przez gracza B;
A, B – wierzchołki, w których decyzje podejmuje odpowiednio gracz A i B.
Źródło: Falk, Fehr, Fischbacher (2003).

* * * * * *

Model oparty na egoistycznych preferencjach stanowi w wielu sytuacjach bardzo dobre narzędzie analityczne. Założenie, że ludzie dbają wyłącznie o swój dochód, jest wygodnym oraz, w wielu przypadkach, nieszkodliwym uproszczeniem. Niewątpliwie ludzie często zachowują się tak, jakby mieli *stricte* egoistyczne preferencje. Są jednak sytuacje, w których stosowanie modelu egoistycznych preferencji może prowadzić nas do bardzo nietrafnych predykcji na temat zachowania decydentów kluczowych z punktu widzenia danego problemu społecznego czy ekonomicznego. E. Fehr i U. Fischbacher (2002) wskazują, że bez uwzględnienia oddziaływania nieegoistycznych preferencji możemy nie być w stanie właściwie badać problemów konkurencji, kooperacji i kształtowania systemów motywacji (ang. *incentives*). Analizując przykładowe zastosowania modeli nieegoistycznych preferencji, autorzy ci przywołują badania pokazujące, że nieegoistyczne preferencje mogą mieć istotny wpływ na takie zagadnienia społeczne i ekonomiczne, jak na przykład kwestia płacenia podatków, redystrybucji, zachowań wyborczych, problem korupcji, niekompletnych kontraktów, optymalnego rozdziału praw własności, optymalnej organizacji struktur przetargowych czy też problem pryncypała-agenta. Ponadto, modele preferencji nieegoistycznych okazują się być również przydatnym narzędziem w analizie takich zjawisk makroekonomicznych, jak np. *uporczywość inflacji* (Driscoll, Holden 2003). We wszystkich wymienionych wyżej przypadkach istotne znaczenia mają interakcje między osobami z różnymi typami preferencji. Rezultatem tych interakcji może być w jednej sytuacji dominacja zachowań egoistycznych, w innej zaś – różnych typów zachowań nieegoistycznych. Modele nieegoistycznych preferencji, dopuszczając możliwość wzajemnych oddziaływań między osobami o heterogenicznych preferencjach, pozwalają na badanie warunków, w jakich dany typ zachowań staje się dominujący. Tym samym pozwalają na określenie, w jakich sytuacjach przyjmowanie założenia o egoizmie graczy jest rzeczywiście uprawnione i nieszkodliwe z punktu widzenia uzyskiwanych wyników teoretycznych, a w jakich stanowi ono istotny hamulec dla rozwoju teorii racjonalnego wyboru, i szerzej, ekonomii czy socjologii.

Bibliografia

Andreoni, James i John Miller. 2002. *Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism.* „Econometrica" 70 (2): 737-753.

Blanco Mariana, Dirk Engelmann i Hans-Theo Normann. 2007. *A Within-Subject Analysis of Other-Regarding Preferences*. Working Paper.

Blount, Sally. 1995. *When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences*. „Organizational Behavior and Human Decision Processes" 63 (2): 131-144.

Bolton, Gary E. 1991. *A Comparative Model of Bargaining: Theory and Evidence.* „The American Economic Review" 81 (5): 1096-1136.

Bolton, Gary E. i Axel Ockenfels. 2000. *ERC: A Theory of Equity, Reciprocity, and Competition*. „American Economic Review" 90 (1): 166-193.

Bolton, Gary E. i Axel Ockenfels. 2005. *A stress test of fairness measures in models of social utility*. „Economic Theory" 25: 957-982.

Brosig Jeanette, Thomas Riechmann i Joachim Weimann. 2007. *Selfish in the End? An Investigation of Consistency and Stability of Individual Behavior.* MPRA Paper 2035. University Library of Munich.

Camerer, Collin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Russell Sage Foundation, New York, New York/Princeton University Press, Princeton, New Jersey.

Charness, Gary. 1996. *Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation*. Economics Working Papers 283. Universitat Pompeu Fabra: Department of Economics and Business.

Charness, Gary i Matthew Rabin. 2002. *Understanding Social Preferences With Simple Tests*. „The Quarterly Journal of Economics" 117 (3): 817-869.

Cox, James C., Daniel Friedman i Steven Gjerstad. 2007. *A tractable model of reciprocity and fairness*. „Games and Economic Behavior" 59 (1): 17-45.

Cox, James C. i Vjollca Sadiraj. 2006. *Direct Tests of Models of Social Preferences and a New Model*. Georgia State University working paper.

Czarnik, Szymon. 2007. *Gry eksperymentalne i manowce racjonalistycznego egoizmu.* Decyzje 8: 27-52.

Driscoll, John C. i Steinar Holden. 2003. *Inflation Persistence and Relative Contracting*. FEDS Working Paper No. 2003-29.

Dufwenberg, Martin i Georg Kirchsteiger. 2004. *A theory of sequential reciprocity*. „Games and Economic Behavior" 47 (2): 268-298.

Engelmann, Dirk i Martin Strobel. 2004. *Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments*. „American Economic Review" 94 (4): 857-869.

Erlei, Mathias. 2003. *Heterogeneous Social Preferences*. TUC Working Papers in Economics 0001. Technische Universität Clausthal: Abteilung für Volkswirtschaftslehre.

Falk, Armin, Ernst Fehr i Urs Fischbacher. 2003. *On the Nature of Fair Behavior*. „Economic Inquiry" 41 (1): 20-26.

Falk, Armin Ernst Fehr i Urs Fischbacher. 2008. *Testing theories of fairness – Intentions matter*. „Games and Economic Behavior" 62 (1): 287-303.

Falk, Armin i Urs Fischbacher. 2006. *A theory of reciprocity*. „Games and Economic Behavior" 54 (2): 293-315.

Fehr, Ernst i Urs Fischbacher. 2002. *Why Social Preferences Matter – the Impact of Non-Selfish Motives on Competition, Cooperation and Incentives*. „The Economic Journal" 112: C1-C33.

Fehr, Ernst i Urs Fischbacher. 2004. *Third Party Punishment and Social Norms*. „Evolution and Human Behavior" 25: 63–87.

Fehr, Ernst i Simon Gaechter. 2000. *Fairness and Retaliation: The Economics of Reciprocity*. „Journal of Economic Perspectives" 14 (3): 159-181.

Fehr, Ernst i Klaus M. Schmidt. 1999. *A Theory Of Fairness, Competition and Cooperation*. „The Quarterly Journal of Economics" 114 (3): 817-868.

Fehr, Ernst i Klaus M. Schmidt. 2006. *A Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories*. W: S. Kolm i J. M. Ythier (red.) „Handbook of the Economics of Giving, Altruism and Reciprocity". North-Holland, s. 615-691.

Geanakoplos, John, David Pearce i Ennio Stacchetti. 1989. *Psychological Games and Sequential Rationality*. „Games and Economic Behavior" 1: 60-79.

Konow, James. 2000. *Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions*.

„American Economic Review" 90 (4): 1072-1091.

Kopczewski Tomasz i Marcin Malawski. 2007. *Ekonomia eksperymentalna: wprowadzenie i najnowsze badania*. „Decyzje" 8: 79-100.

Levine David K. 1998. *Modeling Altruism and Spitefulness in Experiment*. „Review of Economic Dynamics" 1 (3): 593-622.

Malawski, Marcin, Andrzej Wieczorek i Honorata Sosnowska. 2004. *Konkurencja i kooperacja: teoria gier w ekonomii i naukach społecznych*. Warszawa: Wydawnictwo Naukowe PWN.

Rabin, Matthew. 1993. *Incorporating Fairness into Game Theory and Economics*. „American Economic Review" 83 (5): 1281-1302.

Sobel, Joel. 2005. *Interdependent Preferences and Reciprocity*. „Journal of Economic Literature" 43: 392-436.