

# Magdalena Adamus

---

## The Economics of Altruism: On the Rationality of Prosocial Behaviour

---

Diametros nr 57, 1-22

---

2018

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

## **EKONOMIA ALTRUIZMU – O RACJONALNOŚCI ZACHOWAŃ PROSPOŁECZNYCH**

- Magdalena Adamus -

**Abstrakt.** Artykuł prezentuje rozważania dotyczące altruizmu i zachowań prospołecznych, sformułowane na bazie doświadczeń z grą ultimatum. W pierwszej części omówione zostaną związki pomiędzy rozumieniem teorii użyteczności oczekiwanej oraz właściwości człowieka ekonomicznego a współczesnym pojmowaniem altruizmu. Szczególna uwaga zwrócona zostanie na terminologiczne rozróżnienia wskazujące na fakt, że we współczesnej literaturze nie mówimy o jednym, ale o wielu różnych altruizmach. Druga część tekstu obejmuje przegląd wybranych behawioralnych koncepcji wyjaśniających fenomen zachowań prospołecznych. Zaprezentowany zostanie także sposób, w jaki potrzeby, uwarunkowania społeczne, obowiązujące normy i indywidualne czynniki psychologiczne wpływają na skłonność do zachowań altruistycznych.

**Słowa kluczowe:** ekonomia behawioralna, model homo oeconomicus, altruizm, racjonalność, gra ultimatum, gra dyktator, David Gauthier, Daniel Kahneman, Amartya Sen.

Published online: 4 September 2018

### **1. Racjonalny altruizm**

#### **1.1. Uwagi wprowadzające**

Model *homo oeconomicus* jeszcze do niedawna przyjmowany był bezdyskusyjnie jako co najmniej normatywny wzorzec zachowań ekonomicznych. Podmiot w tym modelu charakteryzowany jest za pomocą szeregu cech, o których od dość długiego już czasu wiadomo, że różnią się z właściwościami aktywnych ekonomicznie osób. Jednak zgodnie z postulatem Pareto, tak długo jak teoria pozwala na podawanie skutecznych przewidywań, jej celem nie musi być dążenie do prawidłowego opisywania rzeczywistości. Wśród cech, które stanowią rdzeń *homo oeconomicus* najczęściej wymienia się interes własny, racjonalność, maksymalizację użyteczności oczekiwanej, spójne preferencje, znajomość reguł wnioskowania oraz posiadanie pełnej informacji<sup>1</sup>. Przede wszystkim zaś jego działanie jest celowe, oparte na pełnej wiedzy, a narzędzia służące do realizacji zamierzeń dobiera zawsze w sposób adekwatny<sup>2</sup>.

---

Magdalena Adamus  
Institute of Experimental Psychology  
Centre of Social and Psychological Sciences  
Slovak Academy of Sciences  
Dúbravská cesta 9  
841 14 Bratislava, Slovakia  
e-mail: lena.adamus@gmail.com

<sup>1</sup> Solek (2010): 22.

<sup>2</sup> Szarzec (2014): 195.

Model ten obejmuje zatem szereg założeń, jednak większość z nich spotyka się z otwartą krytyką. Już w latach 50. XX wieku dowiedzieliśmy się, że ludzie nie mają zdolności poznawczych odpowiadających tym opisanym w modelu standardowym. Pojawiły się pierwsze paradoksy (Allais, Ellsberg) jednoznacznie wskazujące, że działanie w warunkach ryzyka i tym samym korzystanie z teorii prawdopodobieństwa nastrocza poważnych, a nawet nierozwiązywalnych trudności. To wtedy Herbert Simon uznał, że ograniczenia poznawcze ograniczają jednocześnie racjonalność i dlatego niezbędny jest nowy, urealniony model człowieka ekonomicznego. Nieco później psychologowie (m.in. Slovic, Lichtenstein, Kahneman i Tversky) dowiedli, że w swoich wyborach ludzie kierują się awersją do ryzyka, a zatem inaczej ważą oczekiwane użyteczności, niż wynikałoby to z teorii standardowej. Jednocześnie wykazali, że założenia psychologiczne można sformalizować i wkomponować w zgrabne i klarowne modele ilościowe<sup>3</sup>. Niedługo po publikacji klasycznego już artykułu Kahnemana i Tverskiego<sup>4</sup> przypuszczony został kolejny atak. Tym razem jednak cios został zadany przez samych ekonomistów: W. Güth, R. Schmittberger i B. Schwarze<sup>5</sup> nieco przypadkiem dowiedli, że ludzie są nie tylko głupszy i bardziej lękliwi, niż homo oeconomicus, ale zarazem mniej samolubni<sup>6</sup>. Skonstruowali oni grę, nazwaną *ultimatum*, która obnażyła słabość kolejnego założenia dotyczącego zachowania człowieka ekonomicznego. Ofiarą tego nie do końca zamierzonego zamachu padł egoizm. Już wcześniej założenie to zostało podane w wątpliwość przez Simona, który bezsprzecznie wykazał, że zawsze gdy maksymalizacja ze względów poznawczych nie jest możliwa, podmiot będzie raczej decydował się na zachowania satysfakcjonujące, czy jakby powiedział G. Akerlof *prawie-racjonalne*<sup>7</sup>. Jednak to rok 1982 można przyjąć jako umowną cezurę, po której nastąpił intensywny wzrost zainteresowania zachowaniami i motywacjami altruistycznymi, a teoria ekonomiczna została wzbogacona o drugi po poznawczym, behawioralny filar wskazujący na szeroko rozumiane czynniki motywacyjne, które systematycznie kierują rozum na manowce irracjonalności.

Ale przecież, jeśli by dobrze przyjrzeć się modelowi homo oeconomicus, to nic w nim nie wskazuje, że człowiek musi kierować się egoizmem. Dlatego w ramach ekonomii behawioralnej bardzo wyraźnie rozdzielono te dwa założenia, racjonalność oraz egoizm, i przyjęto, że zaprzeczenie jednemu nie musi oznaczać z konieczności zanegowania drugiego. Argumentów dostarcza już klasyczne dzieło Adama Smitha, w którym autor wskazuje na rolę czynników społecznych w dokonywanych wyborach, w tym także panującej w danej grupie etyki. Internalizacja obowiązujących norm ogranicza czysty egoizm i wymusza uwzględnienie w swoich wyborach dobra innych osób. Wydaje się, że postulowana przez Smitha norma ma raczej wymiar negatywny: wskazuje granicę, której swoboda jednostki nie powinna przekraczać. Zabraniając krzywdzenia, nie mówi o konieczności czynienia dobra, ono samo wyłącza się zaś jako wypadkowa indywidualnych egoizmów:

---

<sup>3</sup> Mullainathan, Thaler (2000): 1; Rabin (2002): 661.

<sup>4</sup> Kahneman, Tversky (1979).

<sup>5</sup> Güth, Schmittberger, Schwarze (1982).

<sup>6</sup> Solec (2010): 22.

<sup>7</sup> Zob. np. Simon (1955); Consilk (1996); Akerlof (2002).

(...) w pogoni za bogactwem, zaszczytami i wyróżnieniami (człowiek) może uczynić wszystko, wyężyć wszystkie siły, naciągać wszystkie mięśnie, by zdystansować swoich współzawodników. Gdyby jednak miał oszukiwać czy powalić jednego z nich, pobłażanie całkowicie się kończy. Nie mogą pogodzić się z pogwałceniem zasady szlachetnego postępowania<sup>8</sup>.

Ale nieco wcześniej, także w *Teorii uczuć moralnych*, pisał przecież, że: Jakkolwiek samolubnym miałby być człowiek, są niewątpliwie w jego naturze jakieś pierwiastki, które powodują, iż interesuje się losem innych ludzi, i sprawiają, że ich szczęście jest dla niego nieodzowne, choć jedyna przyjemność, jaką może stąd czerpać, to przyjemność oglądania tego<sup>9</sup>.

Ludzie, jego zdaniem, powodowani są namiętnościami, jednak potrafią się samoograniczyć, korzystając z rozumu oraz współczucia wobec innych. I choć zapośredniczona w Stwórcy wiara Smitha w naturę ludzką była nadmierna, to pewne jej echa odnajdujemy w wynikach badań eksperymentalnych, potwierdzających, że en masse ludzie są jednak zdolni do zachowań prospołecznych, może nawet do altruizmu, choć niekoniecznie do bezinteresowności. Pamiętajmy przy tym, że obserwacje Smitha dotyczące motywacji, preferencji czy wyborów w wielu aspektach do dziś pozostają aktualne, choć przecież swoje dzieła stworzył na 100 lat przed powstaniem psychologii w jej nowoczesnej postaci. Chociaż nie mógł posiłkować się zaawansowaną wiedzą na temat struktur psychicznych, a jego język w wielu miejscach wydaje się dziś co najmniej przestarzały, to jednak miał niezwykle trafny zmysł ich antycypowania. Ludzie częściej niż przewiduje ekonomia, płacą podatki, dbają o środowisko, uczestniczą w akcjach charytatywnych czy głosowaniach i zdecydowanie rzadziej „jeżdżą na gapę”. Przed rozstrzygnięciem, czy obserwowane zachowania są altruistyczne czy też nie, skupmy się jednak przez chwilę na eksperymencie, który wywołał całe zamieszanie.

## 1.2. *Gra ultimatum*

Intelektualny zwrot ku badaniu filozoficznego w swej naturze zagadnienia jest tym bardziej zaskakujący, że zastosowane narzędzie bez najmniejszej przesady można określić jako trywialne. A jednak jest to prostota pozorna, ponieważ praktyka obnażyła słabość ortodoksyjnego spojrzenia na proces podejmowania decyzji i to przede wszystkim w warstwie predykcyjnej. Realne zachowanie okazało się odmienne od przewidywanego: gracze proponowali więcej, niż musieli lub nie akceptowali ofert, które uznali za zbyt niskie. Od tamtej pory część ekonomistów zadaje sobie pytanie: czyżby, ku zaskoczeniu wszystkich, konieczne miało stać się uwzględnienie w modelu czynników innych niż egoistyczne?

W zależności od wariantu gracze zostają połączeni w pary przy zachowaniu anonimowości. Pierwszy z graczy, *A*, otrzymuje określoną kwotę pieniędzy, powiedzmy 100 dolarów, i musi podzielić się nią z drugim z graczy. W niektórych postaciach gry dopuszcza się złożenie także „zerowej” oferty. Decyzja drugiego gracza kończy grę. Jeżeli zaakceptował

---

<sup>8</sup> Smith (1989): 123.

<sup>9</sup> Ibidem: 5.

ofertę (Z), cała suma dzielona jest pomiędzy dwóch graczy zgodnie z propozycją gracza A. Jeżeli jednak gracz B odrzuci podział (O), obie strony odchodzą z niczym.

Czy rzeczywiście ta nieskomplikowana gra może stanowić zagadkę? Zastanówmy się nad treścią założenia o racjonalności w jego klasycznej postaci – w tej grze gracz B może zwiększyć swój stan posiadania o wielkość równą (co najmniej) 1. Ponieważ 1 to zawsze więcej niż 0, racjonalny gracz ograniczający się w swoich motywacjach do własnego interesu, zawsze powinien przyjąć ofertę, niezależnie od jej nominalnej wartości. Z drugiej strony racjonalny gracz A zawsze powinien oferować możliwie najmniej – skoro już musi podzielić się pieniędzmi, to najniższa oferta będzie najkorzystniejsza. Jeżeli do tego zakłada, że drugi z graczy jest racjonalny, to nie musi się obawiać, że zakwestionuje zaproponowany podział.

Z perspektywy modelu klasycznego gra wydaje się nieskomplikowana: każdy z graczy dysponuje jedną dominującą strategią i to właśnie ją powinien wybrać. Mimo to spróbujmy na chwilę puścić wodze fantazji. Wyobraźmy sobie kogoś uczestniczącego właśnie w takiej grze jako gracz B. Do podziału jest pula 100 dolarów, a złożona oferta wynosi 1. Jaką podejmie decyzję? Wiedząc coś na temat ekonomii powinniśmy założyć, że jeżeli gracz taki postrzega siebie jako osobę silnie racjonalną i sprawia mu to przyjemność, a do tego zna teorię użyteczności, to zaakceptuje każdą, dowolnie niską ofertę. Tyle teoria. Empiryczne badania prowadzone na całym świecie z dużym poświęceniem i zaangażowaniem pokazują, że rzeczywistość jej przeczy. Przede wszystkim połowa ofert niższych niż 20% puli zostaje odrzucona. Oferty niższe niż 10% zdarzają się sporadycznie, a najczęściej oferowane są kwoty z przedziału 40–50%<sup>10</sup>. Eksperymentalne wyniki wyraźnie kontrastują z założeniem o racjonalności! Jednak samo stwierdzenie faktu to zdecydowanie za mało – potrzebne jest jego wyjaśnienie lub chociaż wrażenie, że do tego wyjaśnienia się zbliżamy. Dlaczego gracze, i to po obu stronach, zachowują się w sposób nieracjonalny? A w szczególności dlaczego ich zachowanie nosi znamiona altruizmu?

### 1.3. *Formy altruizmu*

Argumenty dostarczone przez Smitha w *Teorii uczuć moralnych* nie muszą być przekonujące. Wywiedzione z wątych teologicznych przesłanek przekonania trudno potraktować jako naukowe. Trudniej jednak odrzucić empiryczne dowody potwierdzające, że zachowanie pod względem formalnych charakterystyk może przypominać zachowania altruistyczne. To, czy uznamy je za takie, będzie już jednak miało bezpośredni związek z przyjętą definicją altruizmu. Przyjrzyjmy się zatem definicjom tego pojęcia oferowanym przez słowniki. Słownik języka polskiego lakonicznie stwierdza, że jest to „kierowanie się w swym postępowaniu dobrem innych”, natomiast w drugim znaczeniu jest to „gotowość do poświęceń”. Słowniki języka angielskiego mówią o względzie lub poświęceniu dla dobra innych (Webster), lub skłonności do działań dobrych dla innych nawet ze szkodą dla samego siebie (Cambridge). A zatem definicje słownikowe nie wykluczają, że działania dobre dla innych będą neutralne lub nawet korzystne dla osoby je realizującej. Inaczej będzie jednak jeśli przyjrzymy się znaczeniu

---

<sup>10</sup> Camerer (2003): 49; Dixit, Nalebuff (2009): 72.

terminu altruizm w nauce. Tu okazuje się, że znaczenie przypisywane mu w biologii odbiega od tego powszechnie stosowanego, bliższego zdecydowanie rozumieniu psychologicznemu lub filozoficznemu. W biologii ewolucyjnej jako altruistyczne określa się te działania, które przyczyniają się do zwiększenia dopasowania innych (wyrażonego najczęściej jako liczba potomstwa), zmniejszając jednocześnie własne dopasowanie. A zatem warunkiem koniecznym altruizmu byłoby poświęcenie – w kategoriach bezwzględnych (altruizm silny) lub relatywnie wobec innych (altruizm słaby). Biologiczna definicja nie mówi jednak nic o przyczynach podjętego działania i dlatego takie rozumienie wydaje się sprzeczne ze współczesnymi przekonaniem etycznymi. Jako altruistyczne (a tym samym wartościowane pozytywnie) chcielibyśmy określać takie czyny, które są konsekwencją świadomych intencji, takich jak pragnienie niesienia pomocy.

Teorie doboru krewniaczego, altruizmu odwzajemnionego czy doboru grupowego są jedynie teoretycznymi modelami uprawdopodobniającymi wyewoluowanie i utrwalanie się cechy altruizmu w różnych populacjach. Nie mogą jednak dostarczyć rozstrzygających argumentów. Damasio zauważa, że poza dobrem, które altruści niosą innym, mogą (choć nie muszą) czerpać osobiste korzyści ze swojego zachowania. Do tych korzyści zalicza prestiż społeczny, wyższą samoocenę, uznanie, podziw, a nawet pieniężną gratyfikację. Ostatecznie altruizmowi może towarzyszyć uczucie egzaltacji albo ekstazy<sup>11</sup>. Problem jednak w tym, jak w życiu społecznym określić w sposób jednoznaczny, że dane działanie nie przyniosło żadnej korzyści osobie, która je podjęła? Psychologowie są przygotowani na każdą okoliczność i zawsze mogą uciec się do argumentu z wyrzutów serotoniny, który czyni teorię bezwarunkowego altruizmu bezżębną: wprawdzie pozostaje ona zrozumiałym konceptem, jednak pozbawionym całkowicie praktycznych zastosowań. A zatem co w sytuacji, gdy podmiot działając na rzecz innych odnosi z tego faktu jakies „uboczne korzyści”<sup>12</sup>? Być może bezpieczniej będzie zwrócić uwagę na psychologiczne koncepcje altruizmu, które kładą nacisk na intencje przyczyniające się do podjęcia działania korzystnego dla innych.

Jak się przekonujemy, troska o interes własny nie musi bynajmniej wykluczać nakierowania działań na dobro innych. A zatem, mówiąc umownie, waluta, w jakiej wyrażona jest użyteczność, nie musi sprowadzać się do czysto materialnych wartości (już wspomnieliśmy tu o wyrzutach serotoniny, czyli hormonalnej nagrodzie, która może towarzyszyć działaniom spełniającym definicję altruizmu). Zarówno teorie ekonomiczne, jak i psychologiczne dostarczają argumentów, że działania przynoszące korzyść innym są niejednorodne, a rządzące nimi prawidłowości mogą być pochodnymi bezpośrednio troski o innych, jak i być z nią zupełnie niezwiązane. Tym samym satysfakcja płynąca z formalnie identycznych zachowań może, w zależności od motywacji, być różnie oceniana. Pamiętajmy, że w teorii standardowej racjonalność definiowana jest jako realizacja założenia o maksymalizacji użyteczności, a funkcja użyteczności jest przecież jedynie formalną strukturą spełniającą szereg wymagań stawianych słabemu porządkowi. W swojej matematycznej postaci nie mówi ona zatem nic na temat składowych preferencji, które miałyby podlegać uporządkowaniu, choć z całą pewnością dostrzega zarówno techniczną, jak

---

<sup>11</sup> Damasio (1999): 203.

<sup>12</sup> Sachs et al. (2004): 6.

i filozoficzną trudność „przewalutowania” wartości społecznych na materialne. Tak długo jednak, jak aksjomaty teorii użyteczności zostają spełnione, podmiot może swoją funkcję użyteczności wypełniać dowolną treścią. I to właśnie ten argument, a nie teoretyczne rozważania Smitha, wydaje się rozstrzygający. Istnienie racjonalnych altruistów jest logicznie możliwe. Tylko czy na pewno jesteśmy gotowi zgodzić się na takie rozumienie altruizmu, w którym towarzyszą mu ukryte korzyści? Albo raczej, czy jesteśmy w stanie empirycznie oddzielić prawdziwie altruistyczne działania od tych, w których gratyfikacja ma postać wewnętrznej satysfakcji? I czy do czegoś jest to w ogóle potrzebne?

Niejednorodny charakter motywacji kierujących ludzkimi działaniami odkrył już Abraham Maslow. Wykazał on, że struktura potrzeb ma postać zhierarchizowaną, a u jej podstaw znajdują się elementarne potrzeby związane z fizjologią. Im wyżej, tym potrzeby stają się bardziej wysublimowane, kończąc na samorealizacji. Gdzieś po drodze jednak (na trzecim i czwartym szczeblu) znajdujemy tak ważne społecznie potrzeby jak przynależność i uznanie, które są pochodną życia jednostki w społeczeństwie. Zaspokojenie tych potrzeb jest warunkiem szczęścia jednostki, co stanowi przecież najprostszą, Benthamowską definicję użyteczności. Jak już jednak widzieliśmy, funkcja użyteczności jest pojemna i nie musi ograniczać się do wartości materialnych. Tym samym może w sobie mieścić także pewne elementy altruizmu. Amartya Sen idzie jeszcze dalej i stwierdza nawet, że osoby postępujące zgodnie z modelem *homo oeconomicus* są głupcami, bo nie rozumieją emocji i norm społecznych, a nie rozumiejąc ich, nie mogą prawidłowo funkcjonować w społeczeństwie<sup>13</sup>. I rzeczywiście trudno nie przyznać mu racji. Uwzględniając to wszystko (strukturę potrzeb z nadrzędną rolą uznania i przynależności, poznawczą wrażliwość na przekonania innych oraz obowiązujące normy społeczne), wydaje się już jasne, że sprzeczność między racjonalnością a altruizmem jest pozorna<sup>14</sup>. Aby to dobrze zrozumieć, konieczne jest jednak poczynienie pewnych bardziej zaawansowanych rozróżnień terminologicznych.

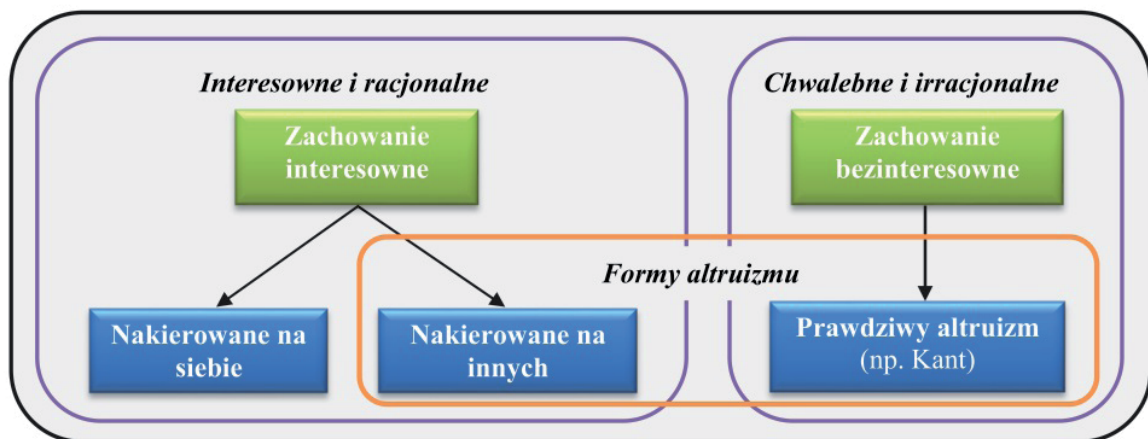
Całe spektrum zachowań, które noszą znamiona altruizmu można podzielić na dwa ogólne typy: interesowne (*self-interested*) i bezinteresowne (*other-interested*), czyli prawdziwie etyczne w sensie nadawanym temu pojęciu np. przez Kanta (por. rysunek 1). Rozumienie jest bliskie temu, jak niekiedy w języku naturalnym definiuje się altruizm. Oceniając dane działanie, bardzo łatwo możemy sprawdzić, czy przyniosło ono korzyści innym, jednak z wielką trudnością albo nawet wcale nie jesteśmy w stanie rozstrzygnąć, czy i jak wielką korzyść osiągnęła osoba, która to działanie podjęła oraz jakie motywy skłoniły ją do jego podjęcia. Korzyści bowiem mogą mieć charakter nieobserwowalny (np. związany z poprawą samopoczucia), po drugie mogą wynikać z egoistycznych motywacji (np. chęci zyskania uznania bądź akceptacji w danej grupie) i po trzecie wreszcie mogą być odłożone w czasie. Nie można się zatem dziwić, że teorie ekonomiczne, które współcześnie odwołują się do altruizmu, koncentrują się raczej na pewnym szczególnym podtypie zachowań interesownych, mianowicie na zachowaniach nakierowanych na innych (*other-regarding*) jako przeciwstawionych zachowaniom nakierowanym na siebie (*self-regarded*). Ponieważ

---

<sup>13</sup> Sen (1977): 336.

<sup>14</sup> Zob. także Burnham (2003).

ekonomia, podobnie jak inne nauki społeczne, pozostaje bezradna w konfrontacji z „prawdziwym” altruizmem, nie ma innej możliwości, niż zająć się analizą faktów obserwowalnych i poddających się kwantyfikacji. Zachowania te w szerokim sensie spełniają definicję altruizmu, choć nie są całkowicie pozbawione pierwiastka egoistycznego (są interesowne) i ostatecznie, właśnie dzięki temu, że przynoszą korzyść innym, zwiększają użyteczność wykonujących je osób. Korzyść własna jest zatem zapośredniczona w korzyści innych osób i wynika faktu udzielenia im pomocy, na przykład ze względu na odczuwane współczucie. Dlaczego zatem uznać w ogóle takie działania za altruistyczne? Być może dlatego, że inaczej pojęcie altruizmu całkowicie straciłoby znaczenie. Przede wszystkim dlatego jednak, że właściwie tylko z takimi niedoskonałymi formami altruizmu mamy na co dzień do czynienia, więc to one mogą stać się przedmiotem badania. Niekiedy jednak, dla uniknięcia terminologicznych nieporozumień, stosuje się określenie zachowania prospołeczne, pamiętając przy tym, że mogą one koncentrować się na użyteczności grupy pośrednio, mianowicie poprzez pomoc jednemu konkretnemu jej członkowi. Taki ogólniejszy termin pozwala uchwycić zarazem kilka zjawisk, które mogą (osobno lub łącznie) wpływać na ostateczny kształt podziału tortu, mianowicie altruizm oraz np. poczucie sprawiedliwości, ale także cały szereg innych motywów, o których będzie mowa w drugiej części tego tekstu.



Rysunek 1: Zależność między zachowaniem altruistycznym i racjonalnym.  
Źródło: Opracowanie własne.

## 2. Altruizm z perspektywy ekonomii behawioralnej

W rozwoju teorii zachowań prospołecznych można wydzielić trzy ogólne fazy. (i) Początkowo badano tylko wyniki końcowe. Podobnie jak w standardowym modelu racjonalności na ocenę zachowania wpływ miała wyłącznie indywidualna percepcja uzyskanej użyteczności, a zatem modele te odzwierciedlały przekonanie o niemal nieograniczonej pojemności funkcji użyteczności i tym samym racjonalności zachowania<sup>15</sup>. (ii) W kolejnej fazie rozwoju dostrzeżono jednak, że wynik to nie wszystko,

<sup>15</sup> Np. Bolton, Ockenfels (2000, 2005); Fehr, Schmidt (1999); Rabin (2002); Charness, Rabin (2002).



a zachowanie musi być oceniane w kontekście czynników behawioralnych, takich jak intencje, czyli motywacje skłaniające jednostkę do danego zachowania (złożenia hojnej oferty lub odrzucenia oferty zbyt niskiej)<sup>16</sup>. (iii) W końcu w trzeciej fazie rozwoju (w przeciwieństwie do modeli (i) ilościowych i (ii) opartych na przekonaniach) pojawiają się modele uwzględniające troskę o status społeczny, tożsamość i wizerunek<sup>17</sup>. Podobną propozycję typologii teorii zachowań prospołecznych znajdujemy u Meiera<sup>18</sup>, który dodatkowo wspomina jeszcze o altruizmie odwzajemnionym (*reciprocal altruism*). Opierając się na pracach Bénabou i Tirole<sup>19</sup> jeszcze innego teoretycznego ujęcia dostarcza Ariely<sup>20</sup>. Otóż dzieli on zachowania prospołeczne ze względu na rodzaj motywacji wpływających na podjęcie (bądź nie) danego działania. Do pierwszej grupy zalicza motywacje wewnętrzne, czyli takie, które mają źródło w prywatnych preferencjach danej osoby. Do tej grupy można zaliczyć przede wszystkim teorie typu (i), włączając w to działania, których celem jest wytworzenie pozytywnego obrazu siebie (we własnych oczach). Następną grupę tworzą zachowania, których przyczyną są zewnętrzne zachęty, takie jak ulgi podatkowe lub bezpośrednie dopłaty, np. związane z zakupem hybrydowego auta lub paneli fotowoltaicznych. I wreszcie trzecia grupa zachowań motywowana jest pragnieniem uzyskania lub utrzymania wysokiego statusu w społeczeństwie. Czynniki te można by najkrócej określić jako „reputacyjne”. Nie zapominajmy jednak, że jest to jedynie typologia czynników, które mogą być ze sobą wewnętrznie sprzężone i występować w różnorodnych kombinacjach, wzmacniając i osłabiając się wzajemnie, nie dostarczając jednocześnie listy wyczerpującej. Zanim jednak przystąpimy do analizy tych zaawansowanych i nietrywialnych koncepcji wyjaśniających przyczyny tendencji do podejmowania działań prospołecznych, powinniśmy przyjrzeć się koncepcji podwójnego wyrafinowania zaproponowanej przez R. Thaler.

### 2.1. *Koncepcja podwójnego wyrafinowania*

Gra ultimatum od niemal samego początku stanowiła wyzwanie dla ekonomistów próbujących wyjaśnić, czym kierują się gracze i czy być może jest to altruizm. Przyjmy się zatem relatywnie prostemu wyjaśnieniu uwzględniającemu dwa wymiary wyrafinowania<sup>21</sup>. Jeden z wymiarów obejmuje klasyczne zdolności poznawcze, o których wiemy już, że nawet przy najbardziej wytężonym wysiłku pozostają ograniczone i zawodne. Drugi jednak nakazuje refleksję nad zachowaniem innych, czyli posługując się psychologicznym terminem, wprowadza elementy teorii umysłu lub intencjonalności (manifestującą się w analizie gry ultimatum jako punkt MAO – minimalna akceptowalna oferta).

Zmierzając do wyjaśnienia paradoksalnych konsekwencji gry ultimatum, można argumentować, że gracze nie potrafili przeprowadzić odpowiedniego rozumowania

---

<sup>16</sup> Np. Rabin (1993), (1998); Blount (1995); Falk et al. (2003).

<sup>17</sup> Np. Akerlof, Kranton (2000), (2010); Bénabou, Tirole (2006).

<sup>18</sup> Meier (2007): 53–54.

<sup>19</sup> Bénabou, Tirole (2003, 2006).

<sup>20</sup> Ariely et al. (2009): 544.

<sup>21</sup> Thaler (2000): 135.

i dlatego nie wybrali racjonalnej strategii. Oczywiście nie można z góry wykluczać, że tak się nie dzieje, ale gra wydaje się na tyle prosta, że nie powinna stanowić intelektualnego wyzwania, a zatem wyjaśnienia w duchu ograniczeń poznawczych nie są w jej przypadku satysfakcjonujące. Można też łatwo upewnić się, czy gracze rozumieją sytuację, po prostu wyjaśniając im, co jest racjonalnym działaniem w tej grze. Być może jednak gracz *A* oferuje więcej, ponieważ boi się, że inaczej jego oferta zostanie odrzucona, a on sam zostanie z niczym. Oznaczałoby to jednak, że z góry zakłada, że drugi z graczy racjonalny nie jest. Wyniki badań empirycznych potwierdzają, że obawa ta nie jest bezpodstawna. Dlatego oferty są wyższe niż mogłyby być, a te, które ocenione zostaną jako zbyt niskie, są systematycznie odrzucane.

Do zbadania biologicznych podstaw zachowania w tej grze wykorzystano najnowsze osiągnięcia technologiczne oraz elementy tzw. neuroekonomii. Jedno z badań zakładało wykorzystanie fMRI lub PET do obrazowania aktywności mózgu podczas gry<sup>22</sup>. Wyniki okazały się zaskakujące. W przypadku ofert znacząco niższych niż 50%, uaktywniała się część mózgu odpowiedzialna za złość i obrzydzenie (wyspa). Jeżeli oferta taka mimo to została przyjęta, uaktywniała się kora przedczołowa, czyli część mózgu odpowiedzialna między innymi za świadomą kontrolę zachowania. Można zatem wprowadzić do tej gry linię wyznaczającą wrażliwość gracza *B* na wartość składanej oferty. Wiemy, że zasadniczo linia ta zlokalizowana jest w pobliżu 50%, choć może być różna dla różnych osób. W zależności od tego, po której stronie MAO znajdzie się oferta ( $V(u)$  – wartość ultimatum), gracz przyjmie ją lub odrzuci. Zatem jeżeli podejrzewamy, że oponent nie będzie zachowywał się zgodnie z modelem człowieka ekonomicznego, a mamy silne przesłanki, by sądzić, że stosuje on różne heurystyki i poddaje się działaniu swoich emocji, to lepiej zaoferować więcej niż dopuszczalne minimum. Lepiej znaczy tu zarówno korzystniej, jak i bardziej racjonalnie<sup>23</sup>.

$$(1) (V(u) \geq MAO) \Rightarrow Z$$

lub

$$(2) (V(u) < MAO) \Rightarrow O$$

Irracjonalne z pozoru zachowanie gracza *A* jest zatem konsekwencją jego oczekiwań dotyczących przyszłego zachowania gracza *B* i pozostaje zgodne z wynikami zaprezentowanych badań. Jest również zgodne z codziennymi doświadczeniami i obserwacjami zachowań w rzeczywistych sytuacjach decyzyjnych<sup>24</sup>. Z perspektywy gracza *A* kluczowe jest ustalenie wartości MAO, czyli skłonności do zaakceptowania propozycji, która zależna jest nie tylko od biologii, ale także od warunków społecznych, indywidualnych doświadczeń i przekonań, norm społecznych, ale i wysokości stawki, osobowości i wiedzy<sup>25</sup>. Ustalając wartość MAO gracz może popełnić błąd, jednak nie

<sup>22</sup> Carerer (2003): 67–68; Dixit, Nalebuff (2009): 75.

<sup>23</sup> Zob. Elster (1998); Thaler (2000).

<sup>24</sup> Por. też Camerer (1997): 167–168.

<sup>25</sup> Camerer (2003): 68; Dixit, Nalebuff (2009): 77.

biorąc tego czynnika w ogóle pod uwagę, zachowuje się zupełnie nierozsądnie. Widzimy więc, że oczekiwania dotyczące decyzji pozostałych uczestników gry stanowią drugi z filarów rozumowania i w konsekwencji racjonalności. Jedno jest pewne: racjonalność w postaci proponowanej w modelu homo oeconomicus jest nieobecna lub dochodzi do głosu wyłącznie w szczególnych okolicznościach. Gdyby jednak to wyjaśnienie okazało się najbliższe prawdy, wówczas musielibyśmy pogodzić się z faktem, że za zachowania prospołeczne w tej grze nie odpowiada altruizm, ale raczej wyrafinowane myślenie strategiczne i dość wąsko pojmowany interes własny. W praktyce obserwowalibyśmy zatem troskę o dobra publiczne (takie jak czyste środowisko lub dotacje na rzecz instytucji kultury), za którą jednak stałaby wizja uzyskania dóbr prywatnych: prestiżu, budowania sieci relacji czy nadania imienia jakiejś budowli<sup>26</sup>. Ekonomia behawioralna nie jest bezradna wobec tych wątpliwości i poddała tę hipotezę skutecznemu, choć niezwykle prostemu testowi.

## 2.2. *Aspekt strategiczny i gra dyktator*

Co by się stało, gdyby grę pozbawić jej strategicznego aspektu? Czy dalej gracze byliby skłonni oferować więcej, niż nakazuje rozsądek? Gdyby tak było, wówczas należałoby szukać przyczyn innych niż tylko obawa przed odrzuceniem zbyt niskiej oferty. Testem tych hipotez stała się prosta modyfikacja gry ultimatum, znana jako dyktator. Guala i Mittone<sup>27</sup> uważają nawet, że jest to gra najprostsza z możliwych, tak prosta, że trudno nazywać ją grą. Zaczyna się dokładnie tak samo, jak ultimatum: gracz otrzymuje pewną kwotę do podziału. Drugi gracz jednak nie może tym razem zaproponowanego podziału odrzucić. Zatem gra kończy się w momencie, w którym padnie propozycja, a ponieważ nie może zostać odrzucona, to znika strategiczny charakter podejmowanej decyzji. Dodatkowo gra toczy się z zachowaniem anonimowości, w niektórych wariantach stosuje się też procedurę *double-blind*. Ten prosty test zaproponowany został przez Kahnemana, Knetscha i Thalerą i dostarczył zaskakujących odpowiedzi na stawiane pytania, które nadal, pomimo swojej niejednoznaczności, mogą wprowadzić zwolenników modelu standardowego co najmniej w zakłopotanie (por. niżej). Jednak gra ta pierwotnie miała służyć po prostu jako warunek kontrolny dla badania anomalii w grze ultimatum i była wstępnym etapem do głębszej analizy skłonności do karania zachowania postrzeganego jako niesprawiedliwe<sup>28</sup>.

Grę opracowaną przez Kahnemana, Knetscha i Thalerą dziś określilibyśmy jako mini-dyktator, ponieważ liczba dopuszczalnych alokacji została wcześniej ograniczona – oznacza to, że dyktatorzy nie mieli pełnej dowolności w dysponowaniu powierzonymi im środkami. Do wyboru mieli tylko dwie możliwości: podzielić pulę 20 dolarów po równo lub zostawić 18 dolarów dla siebie i oddać 2. Graczom zagwarantowano anonimowość, a niektórzy z nich po losowaniu mieli otrzymać wynagrodzenie odpowiadające uzyskanym w grze wynikom. Co zaskakujące aż 76% z nich zdecydowało się podzielić pulę na równe części, choć przecież ich oponent nie mógł w żaden sposób wpłynąć na wynik. Efekt ten nie może jednak zostać uznany za całkowicie wiarygodny, ponieważ gra miała ciąg dalszy,

---

<sup>26</sup> Meier (2007): 52.

<sup>27</sup> Guala, Mittone (2010): 578.

<sup>28</sup> Kahneman et al. (1986a, 1986b).

a zatem „cień przyszłości” mógł wpłynąć na realizację strategii korzystniejszej w dłuższej perspektywie – w kolejnym etapie gry gracze zostali bowiem zgrupowani w trzyosobowe zespoły i poinformowano ich, że mogą podzielić po równo 10 dolarów z graczem „sprawiedliwym” (który wcześniej dokonał podziału 10–10) lub 12 dolarów z graczem „niesprawiedliwym” (który wybrał podział 18–2). Oznacza to, że jeżeli gracz zdecydowałby się dzielić pulę z graczem sprawiedliwym, wówczas niesprawiedliwy odchodzi z niczym (i odwrotnie). Badanie pokazało, że sprawiedliwi gracze z pierwszej rundy aż w 88% przypadków zdecydowali się na podział mniejszej kwoty z innym sprawiedliwym graczem. Spośród niesprawiedliwych graczy tylko 31% wolało dzielić mniejszą pulę w rundzie drugiej<sup>29</sup>.

Ogólnie, mimo sprzyjających okoliczności, zaledwie 40% wszystkich dyktatorów zostawia dla siebie całą pulę, a średnia oferta wynosi około 20%<sup>30</sup>. Manipulując jednak warunkami, można skutecznie wpłynąć na hojność badanych i tak: pełna anonimowość i zarazem zmiana społecznego dystansu, definiowanego jako stopień oczekiwanej wzajemności, redukuje skłonność do dzielenia się i aż 60% uczestników zostawia całą sumę dla siebie, a zaledwie 10% decyduje się przekazać więcej niż 30% stawki<sup>31</sup>. Przekazanie zdjęcia beneficjenta sprawia, że 25% dyktatorów decyduje się przekazać połowę puli, ale aż 58% nie daje nic<sup>32</sup>. Kiedy anonimowości towarzyszy poczucie, że dzielone pieniądze zostały zarobione przez uczestników, zaledwie 3–5% z nich decyduje się cokolwiek zaoferować<sup>33</sup>, jeśli jednak wszyscy uczestnicy (nie tylko dyktatorzy) wykonują pracę, ale tylko połowa z nich jest wynagradzana, skłonność do dzielenia się rośnie do 80%<sup>34</sup>, a kiedy wyłącznie bierni gracze przyczyniają się do pomnożenia puli, 20% dyktatorów jest skłonna oddać przy podziale więcej niż połowę<sup>35</sup>. Na wysokość przekazanej kwoty ma też wpływ znajomość beneficjenta – np. możliwość złożenia darowizny na Czerwony Krzyż sprawia, że korzysta z niej 73% uczestników w porównaniu do zaledwie 27% skłonnych ofiarować coś na rzecz anonimowej organizacji<sup>36</sup>. I w końcu uczestnicy chętniej dzielą się pieniędzmi, jeśli wiedzą, że trafią one do osób niezamożnych lub zostaną przeznaczone na lekarstwa<sup>37</sup>. Z drugiej strony jednak niektórzy twierdzą, że gra dyktator nie mierzy ani altruizmu, ani żadnej innej formy zachowania prospołecznego, a obserwowany efekt jest po prostu wymuszony oczekiwaniami eksperymentatora<sup>38</sup>, zaś badani mogą mieć wątpliwości dotyczące wiarygodności warunków eksperymentu (np. dotyczące istnienia beneficjentów)<sup>39</sup>.

---

<sup>29</sup> Por. Eckel, Grossman (1996a).

<sup>30</sup> Camerer (2003): 49; Dixit, Nalebuff (2009): 77; Guala, Mittone (2010): 579.

<sup>31</sup> Hoffman et al. (1996).

<sup>32</sup> Burnham (2003); Burnham et al. (2000).

<sup>33</sup> Cherry et al. (2002).

<sup>34</sup> Mittone, Ploner (2012).

<sup>35</sup> Ruffle (1998).

<sup>36</sup> Eckel, Grossman (1996b).

<sup>37</sup> Brañas-Garza (2006); Aguiar et al. (2008).

<sup>38</sup> Frohlich et al. (2001); Zizzo (2011).

<sup>39</sup> Wszystkie przytoczone w tym akapicie wyniki pochodzą z różnych replikacji gry „dyktator”. Sama gra za każdym razem formułowana była w ten sam sposób: gracz (dyktator) otrzymywał określoną kwotę do podziału i mógł zdecydować, jaką jej część oddać drugiemu z graczy. Eksperymentalna manipulacja polegała na zmianie kontekstu, w jakim decyzja jest podejmowana, np. poprzez wprowadzenie poczucia, że dzielone pieniądze zostały przez dyktatora zarobione. Dzięki takiemu podejściu możliwe jest zbadanie, w jaki sposób różne zmienne niezależne wpływają na skłonność do zachowań prospołecznych.

Pozytywne oferty w żadnym wariantcie nie zostały jednak całkowicie wyeliminowane, choć niewątpliwie w odpowiednich warunkach znajdujemy się blisko przewidywań modelu standardowego. Mimo że w podstawowej wersji tego eksperymentu gracze nie proponowali tyle, ile w grze ultimatum, to jednak nadal pozostawali pod wpływem jakichś preferencji pro-społecznych. Wiemy już więc, że aspekt strategiczny ma wpływ, ale nie może być jedyną przyczyną obserwowanego efektu. Nadal nie wiemy jednak, jakie dokładnie motywy skłaniają graczy do dzielenia puli w sposób niekorzystny dla siebie i czy te motywy uprawniają nas do określenia takiego zachowania jako altruistycznego. Pozostaje także pytanie o przyczyny obserwowanej zmienności wyników w zależności od wersji eksperymentu – co wpływa na tak dużą rozbieżność ofert w tej przecież bardzo prostej grze? Czy na pewno jest to poczucie „sprawiedliwości”? Czy nie upraszamy zbyt pochopnie naszej wizji świata, w którym faktyczną rolę odgrywa cały szereg złożonych czynników? Na szczęście i to można sprawdzić eksperymentalnie.

### 2.3. Przegląd teorii zachowań prospołecznych

Jak już zauważyliśmy, punktem centralnym współczesnej ekonomii jest model racjonalnego podmiotu, określanego jako *homo oeconomicus*. Autorstwo tego modelu genetycznie łączy się z Adamem Smithem, natomiast pierwsze użycie pojęcia człowieka ekonomicznego przypisuje się Johnowi Stuartowi Millowi, który nigdy tego terminu, w jego dosłownym brzmieniu, nie stosował. Dość precyzyjnie zdefiniował jednak założenia tego jeszcze nienazwanego modelu człowieka i przypisał mu cztery własności czy też popędy, mianowicie do kumulowania dóbr, luksusu, wolnego czasu oraz płodzenia potomstwa, a dodatkowo awersję do pracy. Określenie człowiek ekonomiczny (ang. *economic man*) pada dopiero w reakcji na wyzutą z moralności, zdaniem jemu współczesnych, wizję przedstawioną w *On the Definition of Political Economy; and on the Method of Investigation Proper to It*<sup>40</sup> i datuje się na rok 1888, czyli wydanie pracy Johna Kellsa Ingrama *A History of Political Economy*<sup>41</sup>. Ingram, podobnie jak wielu ówczesnych myślicieli, koncepcję przedstawioną przez Milla uznał za abstrakcyjną i zrównującą człowieka z zarabiającym pieniądze zwierzęciem. To nie znaczy, że Mill nie dostrzegał psychologicznej złożoności człowieka – przeciwnie, wyraźnie stwierdzał, że w ekonomii od tej złożoności należy abstrahować, by w modelu zachować jedynie te cechy, które z jednej strony dla są dla ekonomii istotne, a z drugiej pozwalają na ekonomiczną dedukcję, czyli zbliżają ekonomię do rangi nauki ścisłej. Natomiast łacińskie określenie *homo oeconomicus* po raz pierwszy pada dopiero w 1906 roku w pracy Wilfriedo Pareto *Manual*<sup>42</sup>.

Mijały jednak dziesięciolecia, a model *homo oeconomicus* ewoluował wraz z rozwijającą się ekonomią. Nie sposób więc nie dostrzec, że mimo podobieństw, współczesny model człowieka ekonomicznego różni się od propozycji Milla, dla którego istotne było to „co” wybiera, podczas gdy dziś ekonomia (w tym także ekonomia behawioralna) skupia się na odpowiedzi na pytanie „jak” i to „jak” kojarzone

---

<sup>40</sup> Mill (1836).

<sup>41</sup> Kells (1888).

<sup>42</sup> Pareto (1906).

jest właśnie z racjonalnością, która tym samym nabiera proceduralnego charakteru<sup>43</sup>. Ostatecznie uznaje się, że człowiek ekonomiczny charakteryzuje się szeregiem cech: jest racjonalny, maksymalizuje użyteczność oczekiwaną, posiada pełną informację i doskonale zdolności analityczne, jego preferencje są dobrze zdefiniowane i spójne w czasie, a w swoich wyborach kieruje się wyłącznie własnym interesem. I to właśnie ta ostatnia właściwość, bliska oryginalnemu rozumieniu Milla, wzbudzała i nadal wzbudza ogromne wątpliwości i intelektualny opór. Trudno jednak, by było inaczej, skoro, jak się wydaje, od czasu do czasu obserwujemy zachowania bezinteresowne i wspaniałomyślne, a nade wszystko chcielibyśmy wierzyć, że ludzie do są takich zachowań zdolni. Sen<sup>44</sup> słusznie więc pyta, dlaczego, wiedząc, że założenie o skrajnym egoizmie jest błędne, pokolenia ekonomistów konsekwentnie je stosują? Myliłby się jednak ten, kto uznałby, że głównym celem Sena jest podważenie zasadności samego założenia – po co miałby to robić, skoro jego fałszywość jest tak jaskrawa? On przyjmuje inną strategię – wychodząc od pracy Edgewortha<sup>45</sup>, który interes własny wskazał jako podstawową zasadę ekonomii i wojny – nie stawia pytania o poprawność tej wyidealizowanej wizji człowieka, ale o poprawność wnioskowania i odpowiedzi udzielanych w oparciu o nią. Jego wniosek jest prosty: nawet przyjmując, że ludzie kierują się własnym interesem, musimy uznać porządek ekonomicznego rozumowania. A główne pytanie, które powinniśmy sobie postawić, brzmi: jak nawet w obliczu indywidualnych egoizmów może wyłonić się powszechne dobro i sprawiedliwość? Na to pytanie od dziesięcioleci starają się odpowiadać filozofowie i ekonomiści, choć z różnym powodzeniem. Przede wszystkim dlatego, że, jak trafnie zaobserwował Sen<sup>46</sup>, egoizm i utylitaryzm nie wyczerpują zbioru możliwych motywacji, jakimi w swoim działaniu kierują się ludzie, także w sferze aktywności ekonomicznej. Ostateczny kształt jego rozważania poświęcone moralności, altruizmowi i sprawiedliwości przyjmują w monumentalnej pracy *The Idea of Justice*<sup>47</sup>, w której sięgając do tradycji ekonomicznej, wskazuje na relację pomiędzy racjonalnością, bezstronnością i altruizmem.

Powróćmy jednak na chwilę do klasycznej ekonomii i pytania, jak dzięki synergii egoizmów możliwa jest powszechna pomyślność. Poruszało ono już Edgewortha, a współcześnie najbardziej nośnych odpowiedzi udziela na nie Rawls. Jak sam przyznał, w swoich pracach opiera się na tradycji kontraktariańskiej i w gruncie rzeczy swoją teorię uznaje za bardzo kantowską<sup>48</sup>. Określa ją jako idealną<sup>49</sup> i tak należałoby ją odczytywać: jako projekt filozoficznego rozwinięcia założeń akceptowanych na gruncie ekonomii, wskazujący na ich nieoczywiste konsekwencje wykraczające poza altruizm czy hedonizm, a czynnikiem, który miałby przemienić egoizm w empatię, miałaby być zasłona niewiedzy<sup>50</sup>. Sen jednak zakwestionowałby zasadność

---

<sup>43</sup> Por. Persky (1995).

<sup>44</sup> Sen (1977): 318.

<sup>45</sup> Edgeworth (1881).

<sup>46</sup> Ibidem: 335.

<sup>47</sup> Sen (2009).

<sup>48</sup> Rawls (2009): 20–21.

<sup>49</sup> M. Kuniński (2010) zwraca także uwagę, że teoria Rawls zawarta tym razem w *Prawie ludów* charakteryzuje się realistyczną utopijnością, a jej celem jest rozszerzanie granic tego, co możliwe do osiągnięcia w ramach porządku społeczno-politycznego.

<sup>50</sup> Rawls (2009): 327 oraz Zabdyr-Jamróż (2015): 79.

budowania nowego ładu społecznego w oparciu o empatię. Wprowadził bowiem proste rozróżnienie, znane także w psychologii, zgodnie z którym działania empatyczne wpływają także na dobrostan dobroczyńcy, a zatem nie mogą być traktowane jako prawdziwie altruistyczne. W pełni etyczne, zdaniem Sena, mogą być dopiero te działania, które motywowane są oddaniem lub poświęceniem, i jako takie są nie tylko bezinteresowne, ale także wynikają z silnego przywiązania do prawa i wynikających z niego powinności<sup>51</sup>. Rawls jednak przede wszystkim stawia pytanie o sprawiedliwą wizję społeczeństwa. Sprawiedliwość w takim znaczeniu, jakie nadaje jej ten filozof<sup>52</sup> nakazuje dokonywać oceny względem tych, którym jest najtrudniej, którzy w najmniejszym stopniu korzystają z redystrybucji lub wręcz są nią pokrzywdzeni. Jednak empiryczne testy, choćby te z użyciem gry ultimatum dowodzą, że w praktyce taka wiara w ludzką naturę jest nadmierna. Jasne jest, że w dużej mierze gracze motywowani są interesem własnym, który realizuje się na różne sposoby, a dla którego regularnie poświęcają ideę sprawiedliwego podziału. Zdarzają się i sytuacje, w których egoizm zostaje poświęcony dla sprawiedliwości, jednak zwykle jest to sprawiedliwość Kalego. Dyskomfort, jaki odczuwa gracz w związku z własnymi niskimi wypłatami, nie jest bynajmniej tożsamy z uczuciem, które towarzyszy obserwowaniu prostego faktu, że innym może być gorzej. Nie oznacza to jednak, że dwudziestowieczna filozofia pozostawia nas bezradnymi. Przeciwnie, zarówno Sen, jak i Gauthier czy Nagel dostarczają nam narzędzi do namysłu nad tymi ważkimi kwestiami.

Gauthier stawia pytanie o możliwość racjonalnego altruizmu zakładając, że o ile moralność nie zawsze musi pozostawać w zgodzie z własnym interesem, to z całą pewnością pozostaje w zgodzie z racjonalnością<sup>53</sup>. W innym miejscu zaważa, że być może utożsamienie racjonalnego działania z maksymalizacją nie jest właściwe i to z tego nieporozumienia wynikają trudności współczesnej ekonomii<sup>54</sup>. Krytykując Nagela<sup>55</sup> za podążanie kantowską ścieżką rozumowania, argumentuje, że bezstronności obecnej już w praktycznym rozumie nie należy zakładać, ale dowieść<sup>56</sup>. Jednak wspólne obu filozofom jest uznanie, że wspieranie innych leży bezpośrednio w interesie każdego człowieka<sup>57</sup>. Choć odwołując się do platońskiej *Republiki* Gauthier powtarza, że w naturze człowieka leży nadzieja, że będąc lepszym od innych, uda mu się być niesprawiedliwym, a ten, kto jest prawdziwie człowiekiem zawsze będzie próbował odrzucić ograniczenia moralne<sup>58</sup>. Mimo to w rzeczywistości nie mamy do czynienia z nieustającą wojną wszystkich przeciw wszystkim, istnieje zatem jakaś droga wykroczenia poza wąsko rozumiany własny interes lub przynajmniej jego racjonalnego przededefiniowania. Zgodnie z zaleceniem Gauthiera droga ta wiedzie od bezpośredniej do ograniczonej maksymalizacji, maksymalizacji, która uwzględnia obecność innych i ich

---

<sup>51</sup> Sen (1977): 326.

<sup>52</sup> Rawls (2009).

<sup>53</sup> Gauthier (1986): 1.

<sup>54</sup> Gauthier (1975): 418.

<sup>55</sup> Nagel (1970).

<sup>56</sup> Gauthier (1986): 6.

<sup>57</sup> Nagel (1970): 15.

<sup>58</sup> Gauthier (1986): 19.

wpływ na uzyskiwane przez podmiot wyniki<sup>59</sup>. Ograniczona maksymalizacja, a zatem wybory satysfakcjonujące przeciwstawione wyborom optymalnym, jest także swoistą kompensacją ludzkich słabości czy niedoskonałości. Rozumiejąc więc, że człowiek nie jest samotną wyspą, musimy uznać, że proces racjonalizacji przebiega w interakcji z innymi i musi uwzględniać ich racjonalność, a raczej ograniczenia ich racjonalności. Homo oeconomicus w czystej postaci może być zatem postrzegany jedynie jako głupiec. Nieco wcześniej Sen nazwał go po prostu społecznym idiotą, ponieważ moralność, a w szczególności zachowania prospołeczne, wyłaniają się jako konsekwencje racjonalnego porozumienia równych sobie podmiotów, a to porozumienie musi mieć charakter bezstronny<sup>60</sup>. Bezstronność, podobnie jak u Sena, okazuje się katalizatorem altruizmu wśród racjonalnych podmiotów. Źródła tego rozumowania znajdujemy u Gauthiera już znacznie wcześniej, kiedy odwrócił stawiane przez wielu pytanie i sprawdził, czy możliwy jest racjonalny egoizm<sup>61</sup>. Jego negatywna odpowiedź opiera się na wewnętrznej, logicznej niespójności egoizmu i racjonalności. Ten sposób myślenia zapowiada już wszystko, co później nabrało dojrzałej formy w pracy z 1986 roku. Wniosek jest dość prosty: egoizm jest racjonalny, o ile egoistyczne zachowanie jest możliwe, jednak w otoczeniu innych ludzi, w interakcji z nimi i przy świadomości, że dzisiejsze decyzje rzucają cień na naszą przyszłą reputację, egoizm byłby strategią krótkowzroczną. Podsumowując warto jeszcze raz odwołać się do Nagela<sup>62</sup>, który zamyka swoją pracę konkluzją, że samo stwierdzenie, że ludzie są zdolni do moralności i altruizmu nie oznacza, że z tej zdolności korzystają. Oznacza jednak, że są znacznie bardziej skomplikowani i wielowymiarowi niż wynikałoby to z modelu człowieka ekonomicznego. Czy zatem kształtując ład społeczny powinniśmy zrezygnować z wiedzy, której o człowieku dostarczają ekonomia i psychologia, i trwać na straży nierealistycznych, choć pięknych ideałów? Zwłaszcza jeżeli wiemy, jak w inny sposób osiągnąć równoważne cele, także w sytuacji, gdy norma sprawiedliwości nie jest dostatecznie mocno zinternalizowana, a graczom zależy bardziej na byciu postrzeganym jako sprawiedliwi, niż na samej sprawiedliwości<sup>63</sup>.

Wyciągając wnioski z badania Andreoniego i Bernheima, Ariely wraz z zespołem badawczym postanowili sprawdzić, do jakiego stopnia działanie prospołeczne można przypisać trosce o własny wizerunek i jak wpływają na zachowanie zewnętrzne zachęty<sup>64</sup>. Zgodnie z teorią Bénabou i Tirole ogromne znaczenie dla oceny zachowania ma sposób, w jaki jest ono postrzegane przez innych. Jeżeli społeczność sądzi, że pozornie altruistyczne działanie jest motywowane choćby częściowo interesem własnym, wówczas nie będzie ono przekładało się bezpośrednio na poprawę statusu, ponieważ trudno ocenić, który z czynników miał na nie większy wpływ. A wydaje się, że w tych sprawach społeczeństwo jest skłonne raczej rozstrzygać wątpliwości na niekorzyść „oskarżonego”. Postanowiono zatem sprawdzić, kiedy ludzie chcą

---

<sup>59</sup> Gauthier (1975), (1986).

<sup>60</sup> Sen (1977): 336.

<sup>61</sup> Gauthier (1974).

<sup>62</sup> Nagel (1970): 146.

<sup>63</sup> Por. Bénabou, Tirole (2003, 2006); Andreoni, Bernheim (2009).

<sup>64</sup> Ariely et al. (2009).



postępować właściwie, a kiedy dobrze (głównie dla siebie). Przeprowadzone badanie wymagało dość prostego wysiłku (klikania w klawisze „z” oraz „x”). Im więcej kliknięć wykonała badana osoba, tym większa kwota została przekazana na cel charytatywny. Zaobserwowano, że istotnie wysiłek był tym większy, im bardziej publiczne było to zachowanie. Jednak dodanie zewnętrznej zachęty (finansowej rekompensaty) w warunkach publicznym nie zwiększyło ich zaangażowania (różnica nie jest statystycznie istotna). W warunkach „prywatnych”, kiedy zachęta ani zachowanie nie są znane innym, jej wpływ jest ogromny (nawet dwukrotnie zwiększając wysiłek badanych). Praktyczny wniosek, który płynie z tego badania, jest niezwykle pouczający: jeżeli uczynimy zachowanie publicznym, jako zachęta wystarczą spojrzenia innych. To one posłużą za presję i nagrodę zarazem. Dodawanie finansowych zachęt nie będzie już efektywne. Jednak jeżeli zachowanie jest prywatne, to bez takich zewnętrznych zachęt ludzie najczęściej będą unikać wysiłku i każdego innego kosztu związanego z zachowaniem prospołecznym, i raczej zachowają się egoistycznie (jak w najbardziej anonimowych wariantach gier dyktator i ultimatum). Oczywiście jakaś część badanych zachowywała się prospołecznie (przynajmniej do pewnego stopnia) także w warunkach prywatnym i bez zachęt, jednak średnia liczba kliknięć była o ok. 40% mniejsza (900 kliknięć w porównaniu do 517), co wyraźnie ukazuje, że nie można pokładać nadmiernej nadziei w samym tylko poczuciu sprawiedliwości. Hipoteza, jakoby to ono było głównym (lub nawet jedynym) motorem działania została podana w wątpliwość stosunkowo wcześnie<sup>65</sup>. Z tego dość zagmatwanego badania wypływa jasny wniosek: ponieważ rozkład ofert w grach ultimatum i dyktator jest różny, to samo dążenie do sprawiedliwego podziału nie może stanowić wystarczającego wyjaśnienia. Wśród motywów, które najczęściej wymieniane są jako źródła szeroko rozumianych zachowań prospołecznych najczęściej pojawiają się sprawiedliwość lub proporcjonalność (*equity*), równość (*equality*), wzajemność, współpraca i rywalizacja.

#### **2.4. Kultura, altruizm i najdziwniejsi ludzie świata**

Na jednostkę wpływają, choć zapewne nie z równą siłą, czynniki biologiczne, psychologiczne i kulturowe. Dlatego niektóre rozwiązania oczywiste w jednej kulturze mogą wydawać się nieakceptowalne w innej. A przecież ocena zarówno celu, jak i wykorzystywanych do jego osiągnięcia narzędzi, odbywa się właśnie w odniesieniu do norm. Te zaś przekładają się na motywacje i preferencje. Nie oznacza to bynajmniej, że podmiot nie ma wyboru, przeciwnie, może wybierać spośród wielu wariantów, ale zwykle tych mniej lub bardziej aprobowanych społecznie. Kultura może służyć jako najbardziej ogólny punkt odniesienia – dla większości ludzi zachowanie jest bardziej wartościowe, jeśli jest spójne z oczekiwaniami pozostałych członków grupy i to właśnie dostrzegli w swojej teorii Akerlof i Kranton<sup>66</sup>.

H. Gintis zauważa, nie bez pewnej przesady, że ograniczenia racjonalności mają przede wszystkim źródła społeczne<sup>67</sup>. Oczywiście czynniki te można uznać za szcze-

---

<sup>65</sup> Forsythe et al. (1994): 348.

<sup>66</sup> Akerlof, Kranton (2000, 2010).

<sup>67</sup> Gintis (2009): 249.

gólnego rodzaju presje motywacyjne, jednak badania przeprowadzone wśród przedstawicieli kilkunastu różnych kultur pokazują, że na pewno nie są one bez znaczenia, niezależnie od tego, jak je zaklasyfikujemy. Wszystko zaczęło się od ambitnego doktora, J. Henricha, który przeprowadził badanie za pomocą gry ultimatum wśród Indian w Peru<sup>68</sup>. Średnia oferta wyniosła 26% puli, zaś dominanta zaledwie 15%. Zaskoczony uzyskanymi wynikami uznał, że najwyraźniej popełnił jakiś błąd, którego konsekwencje okazały się bardzo poważne. Wyniki znacząco odbiegały bowiem od tych uzyskiwanych powszechnie w licznych replikacjach gry ultimatum. Różnice były na tyle duże, że postanowił skonsultować się z Boydem i Camererem, którzy polecieli mu przeprowadzić dokładną replikację jego badania na grupie w USA (z zachowaniem proporcjonalności stawki)<sup>69</sup>. Wyniki nie okazały się zaskakujące: spora część badanych oferowała, jak zwykle w takich grach, 50%. Wszyscy trzej zrozumieli wówczas, że Henrich odkrył, że kultura ma znaczenie i to niebagatelne.

Antropologiczne wyjaśnienie tej sytuacji jest stosunkowo proste i bazuje na strukturze społecznej, w której poszczególni członkowie nie są bezpośrednio powiązani z innymi, z wyjątkiem swoich najbliższych krewnych. Nie powstają między nimi bliskie więzi, a szansa, że obce sobie osoby będą musiały ze sobą w przyszłości współpracować, jest niewielka. Jednak wśród wszystkich populacji, które uczestniczyły później w tym badaniu, wyniki okazały się zdecydowanie bardziej zróżnicowane i zaskakujące. Przedstawiciele niektórych grup oferowali często więcej niż połowę (w Paragwaju i Indonezji), ogólnie jednak w 10 z tych grup oferty były istotnie mniejsze niż w standardowych badaniach realizowanych w krajach rozwiniętych. Średnie oferty, w zależności od grupy, wahały się od 25% do 57%, dominujące propozycje mieściły się w przedziale od 15% do 50%, a zatem napotkano społeczności zarówno bardziej, jak i mniej hojne, niż wskazywałyby na to wcześniej publikowane badania. W jednych grupach odmowy praktycznie się nie zdarzały, nawet w przypadku bardzo nierównych ofert, w innych odrzucane były oferty, które przekraczały 50% wartości puli<sup>70</sup>. Henrich nawet posunął się do tego, by nazwać standardowe próby, czyli zwykle studentów amerykańskich uniwersytetów, „najdziwniejszymi na świecie” (WEIRD – **W**estern, **E**ducated, **I**ndustrialized, **R**ich, **D**emocratic)<sup>71</sup>. Mimo to, popadając w błąd fałszywej zgody<sup>72</sup>, jesteśmy gotowi uznać, że inni są do nas podobni. Podczas gdy prawda może być zupełnie inna: nie tylko różnimy się, ale też nie jesteśmy dla ludzkości specjalnie reprezentatywni.

Wnioski płynące z tego zakrojonego na szeroką skalę badania antropologicznego potwierdzają także badania z udziałem dzieci oraz osób autystycznych. Camerer<sup>73</sup> zwraca uwagę na szczególnie kierunek zmiany zachowania u dzieci wraz z wiekiem. Dzieci dojrzewając przechodzą przez trzy fazy rozwojowe (w kontekście altruizmu): początkowo, mniej więcej do piątego roku życia skupiają się wyłącznie na własnych potrzebach, następnie (5–7 lat) dominuje zasada równości, niekiedy doprowadzona

---

<sup>68</sup> Henrich et al. (2001), (2004); Camerer (2003): 68–74.

<sup>69</sup> Camerer (2003): 68.

<sup>70</sup> Por. Henrich et al. (2001, 2004); Camerer (2003): 68–74.

<sup>71</sup> Henrich et al. (2010).

<sup>72</sup> Thaler (2000): 133.

<sup>73</sup> Camerer (2003): 65–67.

wręcz do absurdu (próby dzielenia pojedynczego cukierka na pół), ostatecznie zaczynają rozumieć, że istotna jest sprawiedliwość (*equity*). We wszystkich badaniach przeprowadzonych z udziałem najmłodszych dzieci, wyniki były bardziej zbliżone do modelu homo oeconomicus: to znaczy dzieci nie tylko oferowały niewiele, ale były też bardziej skłonne akceptować niskie oferty (najniższa możliwa oferta była akceptowana w 70% przypadków). Do podobnych wniosków skłania badanie Sally'ego i Hill<sup>74</sup>: autystyczne dzieci są o wiele mniej skłonne do współpracy i zachowań altruistycznych, niż zdrowi dorośli, co jednocześnie współgra z wynikami uzyskiwanymi wśród dzieci zdrowych. Najniższe oferty były równie częste, jak równy podział, podczas gdy w grupie zdrowych równolatków, zerowe oferty w ogóle się nie pojawiały. Z wiekiem sytuacja ta ulega zmianie i zachowanie dorosłych osób z autyzmem coraz bardziej zbliża się do tego obserwowanego w ogólnej populacji. Z jednej więc strony możemy stwierdzić, że wypracowują oni w sobie zdolność do przewidywania zachowań innych (teoria umysłu), z drugiej zaś mogą także przyswajać sobie ogólne normy regulujące zachowanie w ich społeczności. Znając je, uczą się postępować zgodnie z nimi, nawet jeżeli ich nie rozumieją. Podobnie jak przez całe życie uczą się tego także inni ludzie.

### 3. Podsumowanie

Przedstawione tu teorie nie wyczerpują w żadnym razie ogromu tego, co już zostało na temat preferencji społecznych napisane. Z całą pewnością stanowią zaledwie wycinek tego, co zostanie napisane w przyszłości. Teorie zachowań prospołecznych nie powiedziały jeszcze swojego ostatniego słowa, ale to, czego dzięki nim się dowiadujemy okazuje się zarówno interesujące, jak i użyteczne. Wyniki sugerują nam, że zrozumienie fenomenu altruizmu wymaga zdecydowanie większego wysiłku poznawczego, niż pierwotnie zakładała ekonomia i być może także niż zakłada filozofia. Wstępne obserwacje pozwoliły na jednoznaczne stwierdzenie, że funkcja użyteczności musi być pojemniejsza, niż początkowo sądzono, że dwie silne normy (racjonalizm i altruizm) nie muszą się bynajmniej wykluczać. Nie rozstrzygały jednak, co, a w szczególności jak wiele, musi w sobie pomieścić. Wiemy, że ludzie różnią się od siebie i to niekiedy znacząco. Najogólniej można powiedzieć, że około 25% uczestników różnych badań poświęconych zachowaniom prospołecznym zawsze wybiera altruizm, 25% zawsze wybiera egoizm, ale 50% tworzy strefę szarości, w której różne zachowania są możliwe, a na ich kształt wpływ mają okoliczności, w których decyzja jest podejmowana.

Wrażliwość na te okoliczności została odzwierciedlona w licznych wariantach gry ultimatum, które wycinek po wycinku studiowały skłonność do dzielenia się z innymi i próbowały uchwycić w nich pierwiastek bezinteresowności. Wiemy, że w odpowiednich okolicznościach znaczenie może mieć płeć, wiek, wykształcenie, poczucie własności, znajomość beneficjentów, a wreszcie społeczne normy regulujące zasady życia w danej grupie. Nadanie układowi relacji międzyludzkich właściwych instytucjonalnych ram może stać wystarczającą zachętą do tego, by odnaleźć motywację do zachowania zorientowanego między innymi na interes innych<sup>75</sup>. Jedno jest pewne: mimo wszystkich stwier-

---

<sup>74</sup> Sally, Hill (2006).

<sup>75</sup> Por. Fehr, Schmidt (1999).

dzonych podobieństw, altruizm nie jest uniwersalną wartością wynikającą z ludzkiej natury. Być może czysty altruizm (rozumiany jako działanie bezinteresowne i bezwarunkowe) jest czymś wyjątkowo rzadkim, na pewno na tyle rzadkim, że w zasadzie w żadnym eksperymencie nie udało się go potwierdzić. Właściwie każda z przytoczonych teorii pokazuje, w jaki sposób działania prospołeczne przekładają się na indywidualne korzyści. A jednak z różnych powodów ludzie są zdolni do zachowania ze względu na innych, a stojące za ich działaniami motywacje nie powinny być wykorzystywane do umniejszania skali czynionego dobra, nawet jeżeli jest ono tylko warunkowe. Być może postępują tak, by zyskać prestiż, uznanie, wewnętrzną satysfakcję lub by uniknąć ostracyzmu, a być może rzeczywiście kierują się zasadą sprawiedliwości. Dzięki behawioralnym analizom jesteśmy w stanie udzielić przynajmniej częściowej odpowiedzi na te pytania. Jak zauważa Polowczyk, ekonomia behawioralna dzięki uwzględnieniu psychologicznej wiedzy na temat jednostki zwiększa potencjał wyjaśniający ekonomii<sup>76</sup>, a zatem wkracza tam, gdzie klasyczna ekonomia zawodzi. Opinia ta współbrzmi z tą wyrażoną przez Camerera i Loewensteina<sup>77</sup>, mianowicie, że lepsze przewidywania mogą zagwarantować jedynie bardziej realistyczne założenia i to nawet jeżeli płynące z nich wnioski pozostawiają po sobie filozoficzny dyskomfort. Realizm w teorii altruizmu sprowadza się do tego, by uznać, że zachowania prospołeczne mogą mieć różne źródła, ale nie czynią one tych zachowań mniej pożądanymi<sup>78</sup>. Nawet jeżeli te najbardziej wiarygodne teorie zachowań prospołecznych zawierają w sobie skłonność do rywalizowania, to wskazują też na okoliczności, w których możliwa jest współpraca i kiedy znaczenia nabiera sprawiedliwość.

I choć wiemy, zarówno z obserwacji życia codziennego, jak i dzięki wynikom badań eksperymentalnych, że wszystkie te zjawiska występują i mają realny wpływ na kształt społeczeństwa, to jednak najnowsze badania dostarczyły nam przede wszystkim informacji o zakresie i intensywności ich wpływu w konkretnych sytuacjach, a także o tym, jak czynniki sytuacyjne ten wpływ modyfikują. Nie można już zaprzeczyć, że temat ten jest ważnym ogniwem łączącym ekonomię behawioralną i klasyczną, wzbudzającym zainteresowanie po obu stronach, ale jednocześnie niebezpiecznie wkraczającym na obszary wcześniej zarezerwowane dla filozofii czy psychologii. Sukces teorii preferencji społecznych nie zależy jednak jedynie od stwierdzenia odstępstw od modelu standardowego, ale przede wszystkim wiąże się z próbami nadania funkcji użyteczności realnego, matematycznego kształtu, tak aby mogła być ona uniwersalnie stosowana. Wydaje się, że właśnie modele wywodzące się z badania gry ultimatum do pewnego przynajmniej stopnia spełniają te nadzieje. Nie potrafiąc oddzielić działania od motywów, musimy zgodzić się na ocenę tego, co obserwowalne, by dzięki temu możliwie najpełniej zrozumieć preferencje osób dokonujących wyborów. Zrozumienie to nie tylko zaspokaja ciekawość poznawczą, ale być może przede wszystkim przyczynia się do tworzenia świata, w którym najłatwiej ludziom będzie ujawnić to, co w nich najlepsze i tłumić to, co niekoniecznie przynosi im chwałę. Nawet jeżeli sami przy tym mieliby czerpać korzyści. Ich działania mogą być bezwarunkowe, w tym sensie, że nie oczekują wzajemności w przyszłości i nie uzależniają działania od przeszłych zachowań, jednak nie muszą być (i zwykle nie są) bezinteresowne. I choć

---

<sup>76</sup> Polowczyk (2010): 497.

<sup>77</sup> Camerer, Loewenstein (2002): 4.

<sup>78</sup> Por. Rabin (2002).

nadal nie poznaliśmy jednego, pozbawionego skaz równania, w którym zamykałaby się cała wiedza o człowieku, jego wielkość i małość zarazem, to przecież nie znaczy, że wstępując na ten grząski obszar powinniśmy porzucić nadzieję. Możemy oczywiście twierdzić, że taki altruizm to nic innego niż pośredni lub częściowy egoizm, ale dlaczego właściwie nasza szklanka miałaby być zawsze w połowie pusta?

## **Bibliografia**

- Aguiar F., Brañas-Garza P., Miller L.M. (2008), *Moral distance in dictator games*, „*Judgment and Decision Making*” 3 (4): 344–354.
- Akerlof G.A. (2002), *Behavioral macroeconomics and macroeconomic behavior*, „*The American Economic Review*” 92 (3): 411–433.
- Akerlof G.A., Kranton R.E. (2000), *Economics and Identity*, „*The Quarterly Journal of Economics*” 115 (3): 715–753.
- Akerlof G.A., Kranton R.E. (2010), *Identity Economics. How our identities shape our work, wages, and well-being*, Princeton University Press, Princeton.
- Andreoni J., Bernheim B.D. (2009), *Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects*, „*Econometrica*” 77 (5): 1607–1636.
- Ariely D., Bracha A., Meier S. (2009), *Doing Good or Doing Well? Image Motivations and Monetary Incentives in Behaving Prosocially*, „*American Economic Review*” 99 (1): 544–555.
- Bénabou R., Tirole J. (2003), *Intrinsic and Extrinsic Motivation*, „*Review of Economic Studies*” 70: 489–520.
- Bénabou R., Tirole J. (2006), *Incentives and Prosocial Behavior*, „*The American Economic Review*” 96 (5): 1652–1678.
- Blount S. (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 G.E., Ockenfels A. (2000), *ERC: A Theory of Equity, Reciprocity, and Competition*, „*The American Economic Review*” 90 (1): 166–193.
- Bolton G.E., Ockenfels A. (2005), *A stress test of fairness measures in models of social utility*, „*Economic Theory*” 25: 957–982.
- Brañas-Garza P. (2006), *Poverty in dictator games: Awakening solidarity*, „*Journal of Economic Behavior and Organization*” 60: 306–320.
- Burnham T.C. (2003), *Engineering altruism: a theoretical and experimental investigating of anonymity and gift giving*, „*Journal of Economic Behavior and Organization*” 50: 133–144.
- Burnham T.C., McCabe K., Smith V.L. (2000), *Friend-or foe intentionality priming in an extensive form trust game*, „*Journal of Economic Behavior and Organization*” 43: 57–73.
- Camerer C. (1997), *Progress in Behavioral Game Theory*, „*The Journal of Economic Perspectives*” 11 (4): 167–188.
- Camerer C. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press, Princeton.
- Camerer C., Loewenstein G. (2002), *Behavioral Economics: Past, Present, Future*, [w:] *Advances in Behavioral Economics*, C. Camerer, G. Loewenstein, M. Rabin (red.), Princeton University Press, Princeton: 3–51.
- Charness G., Rabin M. (2002), *Understanding social preferences with simple tests*, „*The Quarterly Journal of Economics*” 8: 817–869.
- Cherry T.L., Frykblom P., Shogren J.F. (2002), *Hardnose the Dictator*, „*The American Economic Review*” 92 (4): 1218–1221.

- Consilk J. (1996), *Why Bounded Rationality?*, „Journal of Economic Literature” 34 (2): 669–700.
- Damasio A. (1999), *Błąd Kartezjusza. Emocje, rozum i ludzki mózg*, tłum. M. Karpiński, Dom Wydawniczy Rebis, Poznań 1999.
- Dixit A.K, Nalebuff B.J. (2009), *Sztuka strategii. Teoria gier w biznesie i życiu prywatnym*, tłum. D. Gasper, MT Biznes, Warszawa.
- Eckel C.C., Grossman P.J. (1996a), *The relative of fairness: gender differences in punishment game*, „Journal of Economic Behavior and Organization” 30: 143–158.
- Eckel C.C., Grossman P.J. (1996b), *Altruism in Anonymous Dictator Games*, „Games and Economic Behavior” 16 (2): 181–191.
- Edgeworth F.Y. (1881), *Mathematical Psychics. An Essay on the Application of Mathematics to the Moral Sciences*, C. Kegan Paul & Co., Londyn.
- Elster J. (1998), *Emotions and Economic Theory*, „Journal of Economic Literature” 36 (1): 47–74.
- Falk A., Fehr E., Fischbacher U. (2003), *On the nature of fair behavior*, „Economic Inquiry” 41 (1): 20–26.
- Fehr E., Schmidt K.M. (1999), *Theory of Fairness, Competition and Cooperation*, „The Quarterly Journal of Economics” 114 (3): 817–868.
- Forsythe R., Horowitz J.L., Savin N.E. (1994), *Fairness in Simple Bargaining Experiments*, „Games and Economic Behavior” 6 (3): 347–369.
- Frohlich N., Oppenheimer J., Moore J.B. (2001), *Some doubts about measuring self-interest using dictator experiments: the costs of anonymity*, „Journal of Economic Behavior and Organization” 46: 271–290.
- Gauthier D. (1974), *The Impossibility of Rational Egoism*, „The Journal of Philosophy” 71 (14): 439–456.
- Gauthier D. (1975), *Reason and Maximization*, „Canadian Journal of Philosophy” 4 (3): 411–433.
- Gauthier D. (1986), *Morals by Agreement*, Oxford University Press, New York.
- Gintis H. (2009), *The Bounds of Reason. Game Theory and the Unification of the Behavioral Sciences*, Princeton University Press, Princeton 2009.
- Guala F., Mittone L. (2010), *Paradigmatic experiments: The Dictator Game*, „The Journal of Socio-Economics” 39 (5): 578–584.
- Güth W., Schmittberger R., Schwarze B. (1982), *An experimental analysis of ultimatum bargaining*, „Journal of Economic Behavior and Organization” 3 (4): 367–388.
- Henrich J., Boyd R., Bowles S., Camerer C., Fehr E., Gintis H., McElreath R. (2001), *In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies*, „American Economic Review” 91 (2): 73–78.
- Henrich J., Boyd R., Bowles S., Camerer C., Fehr E., Gintis H. (2004), *Foundations of Human Sociality. Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, Oxford University Press, Nowy Jork.
- Henrich J., Heine S.J., Norenzayan A. (2010), *The weirdest people in the world?*, „Behavioral and Brain Sciences” 33 (2–3): 61–135.
- Hoffman E., McCabe K., Smith V.L. (1996), *Social Distance and Other-Regarding Behavior in Dictator Games*, „The American Economic Review” 86 (3): 653–660.
- Ingram J.K. (1888), *A History of Political Economy*, Adam & Charles Black, London.
- Kahneman D., Knetsch J.L., Thaler R.H. (1986a), *Fairness and the Assumptions of Economics*, „Journal of Business” 59 (4): 285–300.
- Kahneman D., Knetsch J.L., Thaler R.H. (1986b), *Fairness as a Constraint on Profit Seeking: Entitlements in the Market*, „The American Economic Review” 76 (4): 728–741.
- Kahneman D., Tversky A. (1979), *Prospect Theory: An Analysis of Decision under Risk*, „Econometrica” 47 (2): 263–291.

- Kuniński M. (2010), *Sprawiedliwość globalna: szlachetny ideał czy rzetelna teoria?*, „Diametros” 26: 136–153.
- Meier S. (2007), *A Survey of Economics Theories and Field Evidence on Pro-Social Behavior*, [w:] *Economics and Psychology. A Promising New Cross-Disciplinary Field*, B.S. Frey, A. Stutzer (red.), The MIT Press, Cambridge (MA): 51–88.
- Mill J.S. (1836), *On the Definition of Political Economy; and on the Method of Investigation Proper to It*, London and Westminster Review.
- Mittone L., Ploner M. (2012), *Asset Legitimacy and Distributive Justice in the Dictator Game: An Experimental Analysis*, „Journal of Behavioral Decision Making” 25 (2): 135–142.
- Mullainathan S., Thaler R.H. (2000), *Behavioral economics*, „NBER Working Paper Series” 7948, The National Bureau of Economic Research: 1–13.
- Nagel T. (1970), *The Possibility of Altruism*, Princeton University Press, Princeton.
- Persky J. (1995), *Retrospectives: The Ethology of Homo Economicus*, „Journal of Economic Perspective” 9 (2): 221–231.
- Polowczyk J. (2010), *Elementy ekonomii behawioralnej w dziełach Adama Smitha*, „Ekonomista” 4: 493–522.
- Rabin M. (1993), *Incorporating Fairness into Game Theory and Economics*, „The American Economic Review” 83 (5): 1281–1302.
- Rabin M. (1998), *Psychology and Economics*, „Journal of Economic Literature” 36 (1): 11–46.
- Rabin M. (2002), *A perspective on psychology and economics*, „European Economic Review” 46 (4–5): 657–685.
- Rawls J. (2009), *Teoria sprawiedliwości*, tłum. M. Panufnik, J. Pasek, A. Romaniuk, Państwowe Wydawnictwo Naukowe, Warszawa.
- Ruffle B.J. (1998), *More Is Better, But Fair Is Fair: Tipping in Dictator and Ultimatum Games*, „Games and Economic Behavior” 23 (2): 247–265.
- Sachs J.L., Mueller U.G., Wilcox T.P., Bull J.J. (2004), *The evolution of cooperation*, „The Quarterly Review of Biology” 79 (2): 135–160.
- Sally D., Hill E. (2006), *The development of interpersonal strategy: Autism, theory of mind, cooperation and fairness*, „Journal of Economic Psychology” 27 (1): 73–97.
- Sen A. (1977), *Rational Fools: A Critique of the Behavioral Foundations of Economic Theory*, „Philosophy and Public Affairs” 6 (4): 317–344.
- Sen A. (2009), *The Idea of Justice*, The Belknap Press of Harvard University Press, Cambridge, Massachusetts.
- Simon H. (1955), *A Behavioral Model of Rational Choice*, „The Quarterly Journal of Economics” 69 (1): 99–118.
- Smith A. (1989), *Teoria uczuć moralnych*, tłum. D. Petsch, Państwowe Wydawnictwo Naukowe, Warszawa.
- Solek A. (2010), *Ekonomia behawioralna a ekonomia neoklasyczna*, „Zeszyty Naukowe Polskiego Towarzystwa Ekonomicznego” 8: 21–34.
- Szarzec K. (2014), *Model homo oeconomicus w teorii makroekonomii głównego nurtu*, [w:] *Dokona- nia współczesnej myśli ekonomicznej: racjonalność, efektywność, etyka*, U. Zagóra-Jonszta, R. Pęciak (red.), Wydawnictwo EU w Katowicach, Katowice: 195–204.
- Thaler R.H. (2000), *From Homo Economicus to Homo Sapiens*, „Journal of Economic Perspectives” 14 (1): 133–141.
- Zabdyr-Jamróz M. (2015), *The Veil of Ignorance and Solidarity in Healthcare: Finding Compassion in the Original Position*, „Diametros” 43: 79–95.
- Zizzo D.J. (2011), *Do Dictator Games Measure Altruism?*, „CBESS Discussion Paper” 11–15: 1–9.