

Autoselekcja w internetowym eksperymencie *Keynesian p-Beauty Contest*

Tomasz Kopczewski, dr, adiunkt, Wydział Nauk Ekonomicznych, Uniwersytet Warszawski

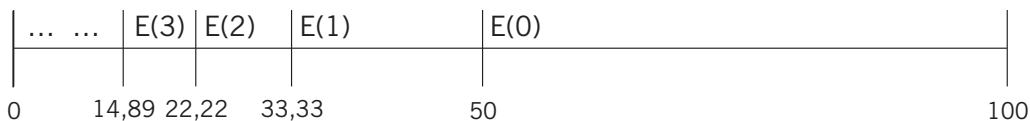
1. Geneza i wykorzystanie eksperymentu

Eksperyment *Keynesian p-Beauty Contest* zawdzięcza swoją bardzo oryginalną nazwę metaforze Keynesa [1936], który porównał zachowanie się inwestorów giełdowych do zachowań uczestników gazetowych konkursów piękności:

...profesjonalne inwestycje można porównać do konkursów gazetowych, w których zawodnicy muszą wybrać sześć najpiękniejszych twarzy ze stu fotografii, zaś nagroda jest przyznawana zawodnikowi, którego wybór jest najbardziej zbliżony do średnich preferencji wszystkich konkurentów; w ten sposób każdy zawodnik musi wybrać nie te twarze, które on sam uzna za najładniejsze, ale te, które jego zdaniem prawdopodobnie przyciągną uwagę innych konkurentów, z których wszyscy patrzą na problem z tego samego punktu widzenia. To nie jest kwestia wyboru tych, które według najlepszych werdyktów są naprawdę najpiękniejsze, ani nawet tych, które przez opinię publiczną rzeczywiście uważane są za najładniejsze. Doszliśmy do trzeciego stopnia, gdzie poświęcamy naszą inteligencję przewidywaniom czego opinia publiczna oczekuje od opinii publicznej...

Kilkadziesiąt lat później Nagel [1995] zauważyła podobieństwo stworzonego przez siebie eksperymentu do opisywanego przez Keynesa zjawiska i ostatecznie eksperyment ten przyjął nazwę *Keynesian p-Beauty Contest*. W tym eksperymencie N uczestników typuje liczbę rzeczywistą z zadanego przedziału — zwykle jest to przedział $[0, 100]$. Zwycięża ta osoba, której podana liczba jest najbliższa wartości $p \cdot \text{średnia ze wszystkich podanych liczb}$. Jedynie zwycięzca otrzymuje ustaloną wygraną pieniężną, pozostali gracze nie otrzymują nic. Istnieje kilka odmian tego eksperymentu w zależności od celu badania, zwykle jest to eksperyment składający się z 10 do 12 rund. Parametr p jest określony przez eksperymentatora i przeważnie jest to wartość $2/3$ [Ho, Camerer i Weigelt, 1998]. Popularność tego eksperymentu wynika z możliwości porównania jednoznacznego rozwiązania teoretycznego samej gry (pełna ra-

cyjność) z wyborami dokonywanymi przez uczestników (ograniczona racjonalność). Wybory graczy są nieobciążone innymi czynnikami motywującymi, np. wzajemnością lub poczuciem sprawiedliwości. Rozwiązanie teoretyczne możliwe jest do uzyskania zarówno przez iteracyjne znajdowanie najlepszych odpowiedzi, jak i przez iteracyjne eliminowanie strategii zdominowanych. Na rysunku 1. zostało przedstawione rozwiązanie gry z $p = 2/3$ osiągnięte przez iteracyjne znajdowanie najlepszych odpowiedzi. Wnioskowanie racjonalne stopnia 0.: ludzie nie podejmują się rozumowania i wybierają losowo liczbę w przedziale $[0, 100]$. Wnioskowanie racjonalne stopnia 1.: jeżeli inni gracze wybiorą w sposób losowy, to strategią wygrywającą jest wybrać $2/3 \cdot 50 = 33,33$. Wnioskowanie racjonalne stopnia 2.: jeżeli wszyscy gracze zakończą swoje rozumowanie na stopniu 1., to najlepszą strategią wygrywającą jest wybrać liczbę $(2/3)(2/3) \cdot 50 = 22,22$ etc. Ogólnie dla K -tego stopnia racjonalnego rozumowania strategią wygrywającą jest wybrać liczbę $(2/3)^K \cdot 50$. Osoba racjonalna powinna dokonać nieskończonej liczby iteracji tego procesu i wybrać 0. Wybór innej liczby oznacza odchylenie się od stanu racjonalności — racjonalność ograniczoną.



Rys. 1.

Stopnie racjonalnego rozumowania w eksperymencie *Keynesian p-Beauty Contest*

Źródło: [Ho, Camerer i Weigelt, 1998].

Na podstawie uzyskanych danych eksperymentalnych można estymować stopień racjonalnego rozumowania [Nagel, 1995]. Przy założeniu, że gracze charakteryzują się adaptacyjnymi oczekiwaniami i prognozują średnią w danej rundzie na podstawie średniej uzyskanej w poprzedniej rundzie, wybór gracza i w rundzie t można przybliżyć wzorem:

$$g_i(t) = m(t - 1)p^K \tag{1}$$

gdzie $g_i(t)$ oznacza wybór gracza i w rundzie t , $m(t - 1)$ średnią z rundy $t - 1$, a p jest parametrem ustalonym przez eksperymentatora $0 < p < 1$. Stopień racjonalnego rozumowania i -tego gracza uzyskuje się przez rozwiązanie równania (1) ze względu na K . Dla rundy pierwszej, gdy gracze nie mają punktu odniesienia, przyjmuje się, że gracze będą antycypowali losowy wybór, stąd też aproksymacja średniej w $t = 0$ jest równa 50. Ujęcie dynamiczne eksperymentu i porównanie wyborów graczy w kolejnych rundach pozwala na analizę procesu uczenia się Nagel [1995].

Eksperyment ten często wykorzystywany jest w studiach komparatystycznych porównujących stopnie racjonalnego wnioskowania różnych grup społecznych. Można wyróżnić kilka typów dokonywanych porównań, ale właściwie wszystkie zmierzały do falsyfikacji założenia o homogeniczności społeczeństwa pod względem stopnia racjonalnego rozumowania. Ich celem było określenie czynników wpływających na heterogeniczność. Badania prowadzone w różnych grupach zawodowych pokazały znaczące różnice między nimi [Camerer, 1997]. Chociaż żadna z grup zawodowych nie przekroczyła 4. stopnia racjonalnego wnioskowania, to można zauważyć zróżnicowanie wyników ze względu zarówno na zawód, jak i czynniki kulturowe (por. tabela 1.). W innym badaniu porównano wyniki osób z doświadczeniem i bez doświadczenia. Wyniki wskazały, że w pierwszej rundzie nie ma znaczących różnic między tymi grupami, a dopiero w następnych rundach różnice te były statystycznie znaczące [Ho, Camerer, Weigelt, 1998]. Ciekawe badania dotyczyły wyborów dokonanych indywidualnie oraz w grupach. Okazało się, że grupy podejmują bardziej racjonalne decyzje niż jednostki [Ho, Camerer, Weigelt, 1998]. Prowadzone były też porównania wyników na podstawie danych uzyskanych z konkursów gazetowych [Bosch-Domènech et al., 2002].

Tabela 1.

Wyniki eksperymentu* *Beauty Contest Game* dla różnych grup zawodowych

Grupa	Średnia	Mediana	Odchylenie standard.	Odsetek wybranej wartości 0	Wielkość próby
Zarządzający portfelem inwestycyjnym	24,31	24,35	16,15	0,08	26
Ekonomiści (z doktoratem)	27,44	30,00	18,69	0,13	16
Rada powiernicza Caltech					
Wszyscy	42,62	40,00	23,38	0,03	73
CEO	37,81	36,50	18,92	0,10	20
Studenci					
Caltech	21,88	23,00	10,35	0,07	27
Niemcy	36,73	33,00	20,21	0,03	67
Singapur	46,07	50,00	28,04	0,02	98
UCLA	42,26	40,50	17,95	0,00	28
Wharton	37,92	35,00	18,84	0,00	35
Uczniowie szkół średnich (USA)	32,45	28,00	18,61	0,04	52

* Eksperyment przeprowadzany dla $p = 2/3$, zerowy stopień racjonalności 50, pierwszy 33, drugi 22,22.

Źródło: [Camerer, 1997].

2. Cel badania

Badania przeprowadzone z wykorzystaniem eksperymentu *p-Beauty Contest Game* ujawniły istnienie heterogeniczności społeczeństwa pod względem osiągniętego stopnia racjonalności oraz szybkości uczenia się — konwergencji do rozwiązania racjonalnego. Jednak pomimo tak wielu badań eksperyment ten jest nadal przedmiotem zainteresowania ekonomii eksperymentalnej. Replikowalność wyników eksperymentu dotyczyła porównań grup, które poprzez proces selekcji eksperymentatora były homogeniczne pod względem możliwości racjonalnego wnioskowania. Problemem nadal pozostaje niejednorodność wewnątrz grupy. Jeżeli gracze w danej grupie są niejednorodni i kilkoro z nich będzie grało nieracjonalnie, to racjonalnym posunięciem jest podanie liczby większej niż zero. O ile większej? Można pokazać to na uproszczonym przykładzie. W grupie N osób K z nich jest nieracjonalnych. Osoby te zawsze wytypują wartość środkową jako swoją liczbę — ich stałą strategią będzie $g_hat = 50$. Strategia osób w pełni racjonalnych powinna uwzględniać nieracjonalność innych i wtedy ich najlepsza strategią g^* jest:

$$g^* = \frac{2}{3} \left[\frac{N-K}{N} g^* + \frac{K}{N} g_hat \right] \quad (2)$$

Po rozwiązaniu równania ze względu na g^* uzyskuje się rozwiązanie równowagi Nasha postaci:

$$g^* = \frac{2 \cdot K \cdot g_hat}{N + 2K} \quad (3)$$

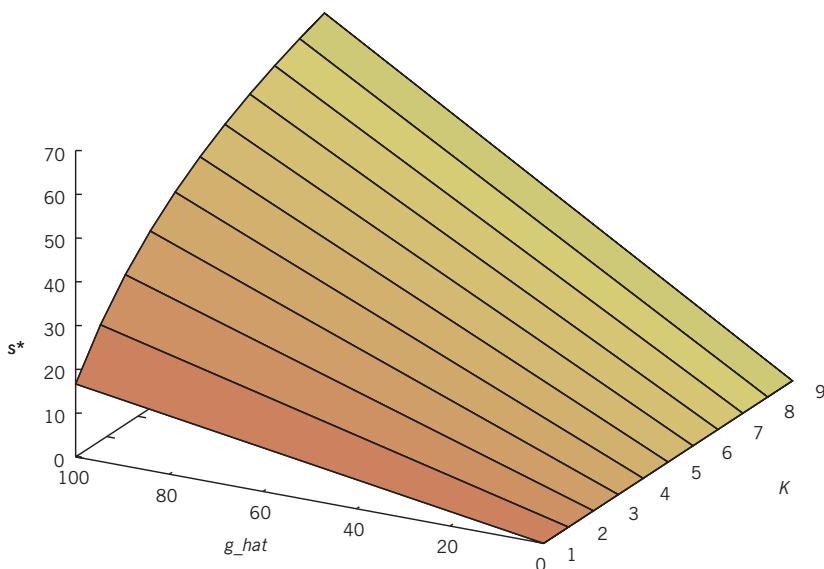
Średnia wskazanych liczb w całej populacji wynosi wtedy:

$$m = \frac{(N-K)}{N} \left[\frac{2 \cdot K \cdot g_hat}{N + 2K} \right] + \frac{K}{N} g_hat \quad (4)$$

Dla populacji $N = 10$, liczby nieracjonalnych graczy $K = 3$ oraz stałej liczby, którą zagrają nieracjonalni gracze $g_hat = 50$, otrzymuje się średnią $m = 28,125$. Oznacza to, że racjonalny gracz powinien zagrać $2/3 \cdot 28,125 = 18,75$. W powyższym przykładzie racjonalni gracze mają pełną informację. Znają zarówno odsetek graczy nieracjonalnych w populacji, jak i wiedzą, jaką liczbę zagrają osoby nieracjonalne. Co się stanie, gdy te założenia zostaną rozluźnione? Nadal wszyscy nieracjonalni typują tą samą liczbę z przedziału $[0, 100]$, ale nie jest znany ich odsetek. Na rysunku 2. pokazano wszystkie możliwe strategie racjonalne g^* dla tego przypadku.

W warunkach eksperymentalnych, gdy w grupie są osoby racjonalne lub znają rozwiązanie tej gry (np. z wykładu teorii gier), to — aby podjąć decyzję racjonalną — muszą uwzględnić proporcje osób nieracjonalnych w populacji oraz stopień ich nieracjonalności, czyli liczbę, którą mogą wybrać, przy czym liczba ta może być różna dla różnych osób. Z punktu widzenia eksperymentatora oznacza to, że na podstawie obserwowanych wyborów nie jest on w stanie określić, z jakimi osobami ma do czynienia, tj. czy z osobami racjonalnymi,

które przeprowadziły powyższe rozumowanie, czy też z osobami, które wybierały liczby na chybił trafił. Problem się jeszcze komplikuje w przypadku dynamicznym i rozgrywaniu kilku rund. Osoby racjonalne muszą estymować na podstawie wyników z poprzednich rund, czy osoby nieracjonalne zmieniają swoje zachowania, czy uczą się i jak szybko. Oczywiście rozróżnienie dychotomiczne na osoby racjonalne i nieracjonalne jest bardzo uproszczone, gdyż osoby te różnią się jedynie stopniem racjonalności, co zostało pokazane w badaniach porównawczych.



Rys. 2.

Strategie równowagowe dla strategii nieracjonalnych graczy z przedziału g_hat [0, 100] oraz liczby nieracjonalnych graczy $K = 1, 2, \dots, 9$

Źródło: opracowanie własne.

Rozwiązaniem tego problemu badawczego było stworzenie modeli, w których zastosowano bardzo restrykcyjne założenia dotyczące procesu wnioskowania. Do tych modeli należy model hierarchii poznawczej (*Cognitive Hierarchy*) [Camerer et al., 2004]. W tym modelu wprowadzono aż siedem założeń upraszczających, aby model był jednoznacznie identyfikowalny. Przyjęto bardzo restrykcyjne założenia, takie jak: a) gracze realizują dyskretną liczbę klas strategii; b) przekonania graczy o względnym udziale uczestników z klas niższych strategii są zgodne z klasą strategii; c) rozkład graczy względem klasy strategii jest rozkładem Poissona. W tym artykule przedstawione będzie skrajnie inne podejście, w którym proces wnioskowania o zachowaniach innych będzie znacznie uproszczony. Przyjęto w nim, że ludzie posługują się najprostszymi heurystykami. Wnioskują na podstawie stereotypów lub auto-stereotypów — „etykietek”, którymi określają innych i siebie. W społeczeń-

stwie polskim istnieją dwie podstawowe etykiety określające ludzi pod względem racjonalnego wnioskowania: „humanisci” oraz „umysły ścisłe”. Podział ten jest nieformalny i jest pochodną podziału przedmiotów szkolnych oraz kierunków studiów. Autor nie spotkał się z udokumentowanymi badaniami społecznymi badającymi ten stereotyp w ujęciu ekonomicznym¹, stąd też, zamiast definiować klasyfikację uczestników opartą na tym schemacie podziału, w badaniu wprowadzony został mechanizm autoselekcji. Przed eksperymentem uczestnicy sami określają, do jakiej grupy należą. Problem autoselekcji wykorzystany był w eksperymencie *Keynesian p-Beauty Contest* [Kocher et al., 2006], w którym uczestnicy decydowali, czy podjąć decyzje indywidualnie, czy jako grupa.

Drugim celem badania było określenie stopnia wpływu środowiska prowadzenia badań na wiarygodność uzyskanych wyników. Autor w swojej pracy koncentruje się głównie na tworzeniu eksperymentów internetowych, w których głównym celem jest edukacja. Jednak na podstawie uzyskanych wyników można wnioskować, że eksperymenty internetowe prowadzone w celach edukacyjnych mogą być pomocnym narzędziem badawczym.

3. Eksperyment

Pierwsze eksperymenty autora uwzględniające proces autoselekcji powstały w celach edukacyjnych, a prowadzone były sieciowo w ramach kursów internetowych oraz stacjonarnie w pracowni komputerowej. Im więcej było przeprowadzonych sesji, tym łatwiej można było zauważyć replikowalność wyników, i to niezależnie od warunków przeprowadzania eksperymentu. Na podstawie uzyskanych doświadczeń autor dostosował kształt eksperymentów tak, aby móc na ich podstawie wnioskować o sile oddziaływania stereotypu na wybory ekonomiczne. Ze względu na środowisko prowadzenia badań wprowadzono następujące zmiany w stosunku do eksperymentów laboratoryjnych:

- Uczestnicy eksperymentu nie wiedzieli, że są badani — eksperyment przeprowadzany był jako dydaktyczny i stanowił zaledwie część gier prowadzonych w czasie zajęć.
- Nie stosowano wynagrodzenia pieniężnego. Wynagrodzeniem było uzyskanie dodatkowych punktów zaliczeniowych.
- W eksperymencie parametr p został zmieniony na $3/4$. Celem tej zmiany było użycie parametru, który rzadko pojawia się w badaniach eksperymentalnych. Studenci mieli dostęp do wyszukiwarek internetowych i w trakcie eksperymentu mogli znaleźć jego opis, co mogło wpływać na przyjętą przez nich strategię.
- Możliwości wyboru zostały zmienione. Studenci mogli samodzielnie wybrać liczbę z przedziału $[0, 100]$ lub mogli wyręczyć się wbudowanym generatorem liczb losowych z tego przedziału, naciskając przycisk LOS. Wpro-

¹ Jedyną pozycją odwołującą się do autostereotypu jest praca Rogaczewskiej i Gołdys [2009], w której analizuje się wpływ autostereotypów Polaków na rozwój ekonomiczny.

wadzenie tej opcji dawało możliwość wyłapania osób, które grały bezwiednie i nie interesował ich eksperyment, a wybór losowy przez komputer był dla nich mniej kosztowny niż wpisanie ręczne liczby. Dzięki temu mechanizmowi osoby bojkotujące eksperyment miały losowy wpływ na wyniki.

- Nie ograniczono możliwości komunikacji między uczestnikami eksperymentu. Jednak, aby mieć kontrolę nad eksperymentem, program miał otwarte okno czatu, pozwalające na w pełni kontrolowane porozumiewanie się — wiadomość do wybranych użytkowników lub do wszystkich.
- Po każdej rundzie podana była liczba zwyciężająca oraz liczba najbliższa wygranej.
- Wprowadzono ograniczenie czasowe 20 sekund na podjęcie decyzji.

Na początku eksperymentu wprowadzono mechanizm autoselekcji. Każdy z uczestników miał zadeklarować, które z określeń lepiej oddaje jego osobowość: „jestem humanistą” / „jestem umysłem ścisłym”. Aby określić wpływ autoselekcji, eksperyment składał się z dwóch tur po 10 rund. W pierwszej turze studenci grali w zadeklarowanych grupach, w drugiej turze wszyscy grali łącznie. Studenci mieli podaną informację, w jakiej grają grupie. Pośrednio mogli uzyskać informację o liczbie graczy w grupie przez komunikację przez czat. W przerwie między turą pierwszą i drugą pojawił się komunikat o połączeniu grup.

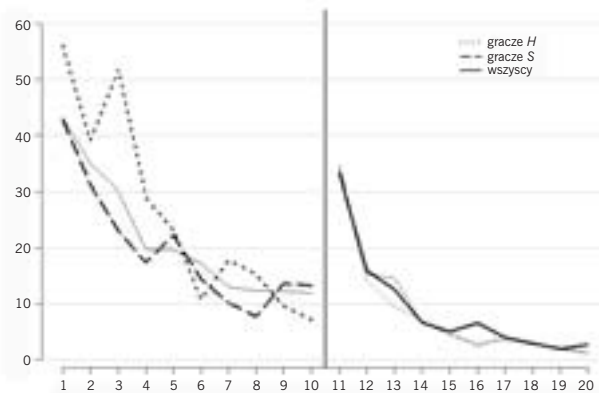
Łącznie zebrano wyniki sześciu grup, przy czym do dalszej analizy użyto wyników z dwóch eksperymentów, w których doszło do podziału na dwie grupy na podstawie deklaracji. W pominiętych w dalszej analizie eksperymentach podział był skrajnie asymetryczny — tylko jedna lub dwie osoby deklarowały przynależność do innej grupy. W przeprowadzeniu wszystkich eksperymentów wykorzystano platformę LabSEE [Kowal, Kopczewski, Borowski, 2008]. Oprócz pozyskania danych ilościowych pozyskane zostały także dane jakościowe. Po eksperymencie jego uczestnicy dyskutowali na czacie. Na podstawie tych rozmów oraz wpisów na formach internetowych autor prowadził zapiski swoich spostrzeżeń. Częściowo badanie przyjęło formę obserwacji uczestniczącej.

4. Wyniki

Na poniższych wykresach oraz tabelach dla eksperymentu 1. i 2. przedstawione zostały wybory uczestników wraz z podstawową informacją dotyczącą głównych charakterystyk przeprowadzonych eksperymentów. Na wykresach wyraźny jest podział na tury: w turze pierwszej jako pogrubione pokazane zostały średnie wartości odpowiedzi z podziałem na grupy humanistów H i umysłów ścisłych S , a jako linia pomocnicza (linia ciągła) średnia dla wszystkich uczestników. W drugiej turze wszyscy gracze zostali połączeni; jako linie pomocnicze zostały zaznaczone średnie wartości grup dla grupy H i S . W tabelach przedstawiono podstawowe statystyki dotyczące poszczególnych grup oraz wartości testu U Manna-Whitneya porównującego rozkłady odpowiedzi w obu grupach.

Eksperyment 1.

Eksperyment przeprowadzony był na grupie studentów kierunku Ekonomiczno-Menedżerskiego WNE UW². Zadeklarowało się 5 osób jako humaniści *H* i 19 osób jako umysły ścisłe *S*.



Rys. 3.

Średnie wartości odpowiedzi z podziałem na grupy humanistów *H* i umysły ścisłe *S*
Źródło: opracowanie własne.

Tabela 2.

Podstawowe statystyki rozkładu odpowiedzi oraz test U Manna-Whitneya

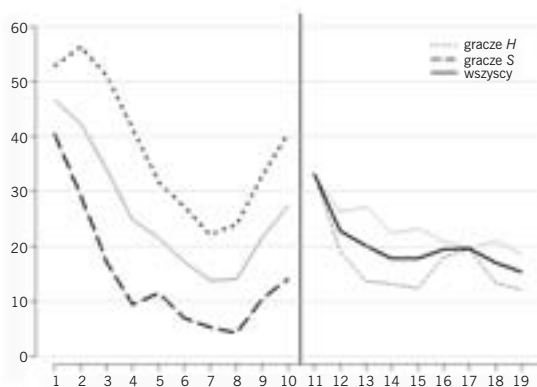
Tura 1. — grupy rozłączne										
Runda	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Średnia (<i>H</i>)	56,00	39,00	51,70	29,00	23,30	10,90	17,80	15,30	9,70	7,10
Odch. std. (<i>H</i>)	18,80	1,40	7,60	11,00	11,00	10,10	3,10	5,50	1,50	1,40
Średnia (<i>S</i>)	42,80	31,20	23,00	17,40	22,30	14,50	10,10	7,80	13,70	13,20
Odch. std. (<i>S</i>)	25,20	20,00	6,50	6,90	31,60	3,70	1,90	3,70	23,90	21,10
Statystyka testu U	24,00	20,00	48,00	37,00	25,00	14,00	34,00	42,00	36,00	25,50
Tura 2. — grupy połączone										
Runda	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Średnia (<i>H</i>)	33,75	14,00	9,60	6,99	4,25	3,05	3,34	2,66	2,20	1,24
Odch. std. (<i>H</i>)	5,30	5,66	2,58	2,00	1,56	1,14	1,93	0,41	1,63	0,26
Średnia (<i>S</i>)	34,59	15,28	14,65	6,52	4,92	2,45	4,16	2,97	1,94	1,18
Odch. std. (<i>S</i>)	18,43	9,78	27,16	2,82	2,89	0,93	1,64	1,35	0,60	0,44
Statystyka testu U	14,50	10,00	39,00	47,00	21,00	44,00	37,00	45,00	53,00	57,50

Źródło: opracowanie własne.

² Wydział Nauk Ekonomicznych, Uniwersytet Warszawski

Eksperyment 2.

Eksperyment przeprowadzony był na uczestnikach kursu przedmiotów ogólnouniwersyteckich COME IBIZA UW. W badanej grupie występowała znaczna niejednorodność ze względu na kierunki studiów. Grupa składała się z studentów matematyki i informatyki (7), psychologii (3) polonistyki i lingwistyki (3) oraz innych kierunków i specjalności. Grupa była niejednorodna pod względem wykształcenia: studia licencjackie, magisterskie oraz doktoranckie.



Rys. 4.

Średnie wartości odpowiedzi z podziałem na grupy humanistów *H* i umysły ścisłe *S*

Źródło: opracowanie własne.

Tabela 3.

Podstawowe statystyki rozkładu odpowiedzi oraz test U Manna-Whitneya

Tura 1. — grupy rozłączne										
Runda	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Średnia (<i>H</i>)	59,1	56,3	51,1	41,5	31,7	27,3	22,2	23,9	32,8	40,6
Odch. std. (<i>H</i>)	25,2	26,4	19,5	10,4	10,4	20,2	11,2	18,7	30,2	31,0
Średnia (<i>S</i>)	39,8	29,1	17,1	9,4	11,4	6,9	5,3	4,3	10,5	14,0
Odch. std. (<i>S</i>)	25,5	12,5	10,0	3,9	25,7	3,3	4,4	2,2	25,8	24,6
Statystyka testu U	118,0	154*	189*	182*	182*	195*	185,5*	195*	182,5*	184*
Tura 2. — grupy połączone										
Runda	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Średnia (<i>H</i>)	47,1	32,9	26,4	27,2	22,5	23,2	20,9	19,7	20,9	18,6
Odch. std. (<i>H</i>)	11,1	5,9	23,7	19,3	17,2	26,6	19,5	11,5	21,5	14,2
Średnia (<i>S</i>)	41,1	33,4	19,1	13,6	13,2	12,3	18,1	19,5	13,2	12,1
Odch. std. (<i>S</i>)	21,6	20,1	3,3	1,8	1,4	3,3	23,6	23,4	2,6	1,8
Statystyka testu U	146,5*	135,5	109,5	155,5*	142*	117,0	121,0	140,0	123,0	122,5

Źródło: opracowanie własne.

Na podstawie analizy powyższych wyników ilościowych można sformułować kilka znaczących wniosków:

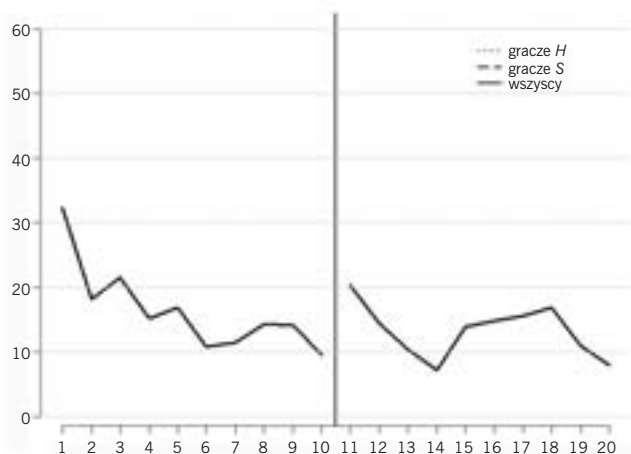
Po pierwsze, zakorzeniony stereotypowy podział na humanistów i umysły ścisłe znalazł swoje odzwierciedlenie w procesie autoselekcji. Kilka grup charakteryzowało się skrajną asymetrią wyborów, ale w żadnym z eksperymentów nie było sytuacji, w której wszyscy zadeklarowali przynależność tylko do jednego typu. Na podstawie rozmów oraz analizy wpisów na forum można stwierdzić, że wśród studentów podział ten funkcjonuje i jest przyczyną tworzenia się antagonizmów między nimi.

Po drugie, mechanizm autoselekcji silnie różnicuje grupy pod względem stopnia racjonalnego wnioskowania pod warunkiem, że grupa jest niejednorodna pod innymi względami. Dla grupy studentów WNE UW (ten sam rocznik, ten sam kierunek) (eksperyment 1.) test U Manna-Whitneya wskazywał, że różnice między grupą humanistów i umysłów ścisłych miały charakter incydentalny. W przypadku grupy niejednorodnej (eksperyment 2.) w turze 1. z podziałem na grupy wartość testu w dziewięciu rundach wskazywała na znaczące różnice między tymi grupami. Jedynie w pierwszej rundzie nie było podstaw do odrzucenia hipotezy o braku różnic w rozkładzie odpowiedzi w tych grupach. Jest to wynik zbieżny z wynikiem Ho, Camerera i Weigelta [1998], gdzie wyniki pierwszej rundy w dwóch grupach z różnym poziomem doświadczenia były zbliżone, a różnice w wyborach pojawiały się dopiero w następnych rundach. Mechanizm autoselekcji różnicuje te grupy pod względem szybkości konwergencji.

Po trzecie, wyniki eksperymentów wskazują na istnienie mocnego efektu połączenia grup, który powoduje cofnięcie się wyników procesu uczenia się. Jest to zaskakujący wynik i świadczy z jednej strony o zaburzeniu procesu uczenia się przez zmianę liczebności grupy, a z drugiej strony jest to sygnał o wpływie autoselekcji. Będąc w grupie umysłów ścisłych, gracze zawyżali swoje deklaracje, w grupie humanistów zaś zaniżali. Po połączeniu efekt ten zanika. Jest to bardzo ciekawy wynik, który powinien być dokładniej zanalizowany w dalszych badaniach eksperymentalnych. O sile procesu autoselekcji świadczą wyniki innych eksperymentów, w których nie dokonał się skrajnie asymetryczny proces podziału na grupy *H* i *S*. W obu turach gracze byli połączeni, gdyż nie było wystarczającej liczby graczy do podziału na grupy *H* i *S*. Informacja ta nie została publicznie podana — gracze myśleli, że w turze pierwszej grają w grupach. Po rzekomym połączeniu ich proces uczenia się został cofnięty (por. rysunek 5. i 6.).

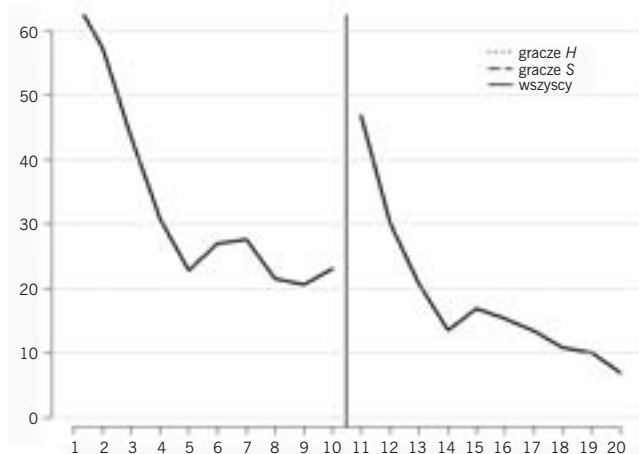
Na podstawie zapisów informacji jakościowej uzyskanej w trakcie eksperymentu można sformułować następujące wnioski:

Po pierwsze, mechanizm autoselekcji był nieco osłabiony przez decyzje strategiczne podjęte podczas deklaracji typu osoby. Kilku uczestników eksperymentu zadeklarowało się jako humaniści, chociaż czuli się umysłami ścisłymi, co wynika z zapisu ich rozmów. Efekt ten był potwierdzony też przez porównanie deklaracji z typem studiów zamieszczonym w ankiecie. Osoby te

**Rys. 5.**

Średnie wartości w eksperymencie ze skrajnym podziałem na grupy — eksperyment prowadzony wśród studentów ekonomii, wszyscy oprócz jednej osoby zadeklarowali się jako umyślnie ściśle.

Źródło: opracowanie własne.

**Rys. 6.**

Średnie wartości w eksperymencie ze skrajnym podziałem na grupy — eksperyment prowadzony wśród studentów nauk humanistycznych — kurs internetowy COME UW, wszyscy oprócz jednej osoby zadeklarowali się jako humaniści

Źródło: opracowanie własne.

uważały, że w grupie humanistów łatwiej będzie im wygrać jako osobom z umiejętnościami matematycznymi.

Po drugie, możliwość komunikacji wpływała na tworzenie się koalicji, które próbowały manipulować wynikami gier. Próby manipulacji odbywały się zwykle od 6.–7. rundy, co jest widoczne na przedstawionych powyżej wykresach jako odwrócenie tendencji i podniesienie się średniej deklarowanej liczby. Tworzone były małe koalicje, w których jedna osoba przyjmowała strategię samobójczą i wybierała 100, a druga osoba wybierała nieznacznie wyższą niż ostatnia wartość wygrywająca, licząc na wzrost przeciętnej. W następnej rundzie uczestnicy albo powtarzali schemat zamieniając się strategiami, albo grali na niższą średnią, jeżeli żaden z nich nie wybierze maksymalnej liczby.

5. Wnioski i dalsze badania

Przedstawione powyżej wyniki badań mają charakter pilotażowy zarówno w obszarze metodologii badań eksperymentalnych, jak i samego problemu racjonalnego wnioskowania i procesu uczenia się. Metodologia badań eksperymentalnych on-line jest wciąż tworzona i nie ma jasnych wytycznych dotyczących warunków, jakie musi spełniać eksperyment internetowy, aby jego wyniki były wiarygodne. W przeciwieństwie do podejścia, w którym postuluje się prowadzenie eksperymentów internetowych tylko w dużej skali³, autor uważa, że specyfika medium, jakim jest Internet, pozwala na włączenie danych jakościowych do analizy, a przez to tworzenie eksperymentów w relatywnie małej skali. Warunkiem pozyskania wiarygodnych wyników jest przede wszystkim stworzenie narzędzi informatycznych, które pozwolą na lepszą kontrolę warunków prowadzenia badań eksperymentalnych. Na podstawie doświadczeń uzyskanych podczas prowadzenia eksperymentu *Beauty Contest* można uznać, że eksperymenty internetowe w małej skali mogą być niezwykle pomocnym narzędziem do tworzenia badań pilotażowych. W eksperymencie laboratoryjnym bardzo ściśle określenie zakresu badania i sposobu jego osiągnięcia powoduje, że zachowania ludzi są zbyt skanalizowane. Eksperyment, który daje większą swobodę graczom, pozwala dostrzec efekty, jakie nigdy nie pojawiły się w laboratorium. Może być to badanie, jak ludzie próbują walczyć z ograniczeniami narzuconymi przez daną sytuację decyzyjną. W analizowanym eksperymencie jest to próba tworzenia koalicji, w której jeden z uczestników poświęca się i wybiera maksymalną liczbę. Dla autora bardzo zaskakującym wynikiem była replikowalność wyników uzyskanych, porównywalna do wyników innych autorów, oraz efekt połączenia. Uczestnicy eksperymentu po połączeniu grup cofali się w procesie uczenia się. Jest to zaskakujące tym bardziej, że rola doświadczenia w podejmowaniu decyzji ekonomicznych jest jednoznaczna. Wynik ten wymaga powtórzenia eksperymentu w warunkach laboratoryjnych, aby — z jednej strony — określić, co go

³ Właściwie nie ma skodyfikowanych wytycznych dotyczących prowadzenia eksperymentów przez Internet. Można oprzeć się jednak na pracach już prowadzonych, np. Centre for Experimental Economics <http://www.econ.ku.dk/cee/>.

determinuje: zwiększenie liczebności próby czy zmiana nastawienia graczy przez myślenie stereotypowe wywołane przez mechanizm autoselekcji, a z drugiej — uwiarygodnić same badanie internetowe w małej skali, dzięki któremu efekt ten został zauważony.

Bibliografia

- Bosch-Domènech A., Montalvo J., Nagel R., Satorra A., 2002, *One, Two, (Three), Infinity, ...: Newspaper and Lab Beauty-Contest Experiments*, „American Economic Review” nr 92, s. 1687–1701.
- Camerer C., 1997, *Progress in Behavioral Game Theory*, „The Journal of Economic Perspectives” Vol. 11, No. 4. (Autumn), s. 167–188.
- Camerer C., Ho T., Chong J., 2004, *A Cognitive Hierarchy Model of Games*, „The Quarterly Journal of Economics” Vol. 119, No. 3 (Aug.), s. 861–898.
- Camerer C., Ho T., Chong J., 2004, *Behavioural Game Theory: Thinking, Learning and Teaching*, w: *Advances in understanding strategic behaviour: game theory, experiments and bounded rationality: essays in honor of Werner Güth*, Palgrave Macmillan, New York.
- Duffy J., Nagel R., 1997, *On the Robustness of Behaviour in Experimental ‘Beauty Contest’ Games*, „Economic Journal” nr 107(445), November, s. 1684–700.
- Ho T., Camerer C., Weigelt K., 1998, *Iterated Dominance and Iterated Best Response in Experimental ‘p-Beauty Contests’*, „The American Economic Review” Vol. 88, No. 4 (Sep., 1998), s. 947–969.
- Keynes J.M., 1936, *The General Theory of Employment, Interest and Money*, Macmillan Cambridge University Press, wyd. pol. *Ogólna teoria zatrudnienia, procentu i pieniądza*, PWN 2003, Warszawa.
- Kocher M., Sutter M., 2005, *The Decision Maker Matters: Individual versus Group Behaviour in Experimental Beauty-Contest Games*, „The Economic Journal” Vol. 115, No. 500 (Jan.), s. 200–223.
- Kocher M., Strauss S., Sutter M., 2006, *Individual or team decision making — Causes and consequences of self-selection*, „Games and Economic Behavior” nr 56, s. 259–270.
- Kowal P., Kopczewski T., Borowski R., 2008, *Oprogramowanie do tworzenia eksperymentów ekonomicznych on-line*, „Ekonomia” nr 21.
- Nagel R., 1995, *Unraveling in Guessing Games: An Experimental Study*, „American Economic Review” nr 85, s. 1313–1326.
- Rogaczewska M., Gołdys A., 2009, *Polacy — w pułapce autostereotypów?*, Instytut Badań nad Gospodarką Rynkową, Gdańsk.

Abstract **Auto-selection in an Online Experiment, p-Beauty Contest Game**

Keynesian p-Beauty Contest (KpBC) is one of the most commonly used laboratory experiments to study the degree of rationality of people and the learning process. It is also applicable in education as a tool to illustrate both the concept of rationality and iterative elimination of dominated strategies. The aim of this paper is to present the results of online teaching experiments in which an auto-selection mechanism has been used based on the existing social stereotypes. The repeatability of the results shows that the experiments are resistant to the distortions arising from the use of the online environment, and that a developed methodology allows to carry out reliable experimental studies.

