

MAGDALENA ADAMUS

(Kraków)

**KONWENCJONALNE UJĘCIE STRATEGII
W ROZSTRZYGANIU PROBLEMÓW
KOORDYNACYJNYCH**

WSTĘP

W tej pracy przedstawiam sposób, w jaki zagadnienie komunikacji dokonującej się za pośrednictwem języka można ująć w kategoriach teorii gier. Pomysł swojej pracy, a także główny tok argumentacji zawdzięczam przede wszystkim Davidowi Lewisowi i jego książce: *Convention*¹ z roku 1969, w której pokazuje, jak dzięki ujęciu porozumienia w terminach teorii gier, a zwłaszcza szczególnego jej fragmentu dotyczącego problemów koordynacji, można uzasadniać tezę konwencjonalizmu językowego. Zanim jednak omówię rdzeń tego tematu, w pierwszej części tej pracy przedstawię ogólne definicje dla teorii gier. Dla potrzeb niniejszego opracowania pomijam rozważania dotyczące definiowania konwencjonalizmu i przyjmuję roboczą definicję samej konwencji.

Definicja 1: Konwencja to zespół uznawanych przez pewną grupę użytkowników danego języka powtarzalnych zachowań, zarówno językowych, jak i pozajęzykowych oraz związanych z nimi procesów rozumienia i przetwarzania informacji (konwencja nie musi być zwerbalizowana, ale każdorazowo, jeśli wystąpi taka potrzeba, musi dać się zwerbalizować).

¹ D. Lewis, *Convention*, Harvard University Press Cambridge, Massachusetts, 1969.

I. TEORIA GIER

I. 1. WSTĘPNE OMÓWIENIE TEORII GIER:
HISTORIA, TEORIA I ZASTOSOWANIA

Teoria gier ma za zadanie badać strategie wybierane przez racjonalnych „graczy” w celu uzyskania preferowanych rezultatów, jest zatem narzędziem opisu ludzkich zachowań. Inaczej mówiąc, dotyczy ona sytuacji, w których racjonalny podmiot działa, uwzględniając możliwe zachowania pozostałych racjonalnych agentów. Została ona sformułowana w 1944 roku przez Johna von Neumanna i Oskara Morgensterna² dla potrzeb ekonomii i nie miała dostatecznie rozbudowanej aparatury matematycznej, stąd jej użycie było obwarowane licznymi dodatkowymi warunkami i tym samym ograniczone do szczególnych przypadków. Po wojnie jednak nastąpił bardzo intensywny rozwój tej teorii, głównie za sprawą Johna Nasha, który w 1994 roku otrzymał za swoje osiągnięcia Nagrodę Nobla.

Współcześnie wskazuje się na historyczne antycypacje tej teorii, w których za każdym razem daje się zauważyć pewien wspólny element, który można by nazwać „logiką sytuacyjną”, wyznaczoną przez preferowane wyniki oraz zbiór potencjalnych przewidywań dotyczących zachowań pozostałych agentów. Kiedy bowiem wchodzimy w interakcje z przedmiotami, istotne dla wyniku działania są nasze intencje i odpowiednio dobrane środki. Jednak w przypadku interakcji z innym racjonalnym agentem konieczne jest branie pod uwagę również jego strategii. Aby to wyjaśnić posłużę się przykładem podanym przez Dona Rossa³.

Załóżmy, że chcę przekroczyć rzekę, przez którą poprowadzone są trzy mosty. Pierwszy jest najzupełniej bezpieczny i nie ma na nim żadnych przeszkód, drugi poprowadzony został poniżej klifu, z którego spadają spore fragmenty skały, na trzecim natomiast są jadowite węże. W takiej prostej sytuacji oczywistym wyborem jest most pierwszy, a kolejność dwóch pozostałych zależy od dodatkowych parametrów takich jak: częstotliwość spadania kamieni, czy ilość węży. Przypuśćmy jednak, że sytuacja jest nieco bardziej skomplikowana i po drugiej stronie rzeki czeka ktoś, komu bardzo zależy na uniemożliwieniu mi przekroczenia rzeki. Założmy też, że jest to osoba bardzo zdespero-

² J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton 1944.

³ Don Ross, University of Cape Town; Stanford Encyclopedia of Philosophy.

wana, a także uzbrojona i niebezpieczna. Oczywiście przeprowadzi ona odpowiednie wnioskowanie: najbezpieczniejszym mostem jest pierwszy, zatem powinna zacząć się właśnie tam. Jeśli jednak ja pomyślę, że tak zrobi, wybiorę drugi w kolejności most, skoro niebezpieczeństwo na nim jest mniejsze, niż w przypadku nadziania się na napastnika. Sytuację tę można komplikować w dowolny, nawet absurdalny sposób, niemniej jednak już teraz widoczne jest, że decyzje, jakie podejmą poszczególni uczestnicy, zależą od ich przewidywań dotyczących zachowań pozostałych uczestników „gry” – zatem muszą oni uwzględnić dodatkowy, niezależny od nich samych czynnik, który umożliwi podjęcie decyzji i wykonanie jakiegokolwiek ruchu. Co więcej, to, co uczestnik wie, jak również to, czego nie wie, znacząco wpływa na analizę sytuacji i dalsze postępowanie.

Teoria gier jest jednak tylko idealizacją realnych zachowań, w związku z tym tylko zwiększanie ilości danych pozwala na skuteczniejsze obliczanie możliwych wyników, nigdy jednak nie pozwala na ich bezbłędne przewidywanie. Dzięki jej aparatowi logicznemu można powiedzieć zatem, co racjonalni gracze powinni zrobić, ale nie co robią. Jest kilka przyczyn niedoskonałości tej analizy. Przede wszystkim jest tak dlatego, że gry rozgrywane regularnie przez rzeczywistych graczy są bardzo skomplikowane i wymagają uwzględnienia wielu zmiennych czynników, które wariują wysokości wypłat. To, co obecnie potrafimy, to wybór „czynników istotnych” i częściowa analiza sytuacji. Po drugie, założenie o racjonalności zachowań poszczególnych graczy często jest podważane z powodu działania pozornie wymykającego się takiemu rozumieniu. To oczywiście nie wyklucza, że zachowanie uważane za irracjonalne jest w rzeczywistości racjonalne w obrębie szerszej gry. Najpoważniejszym jednak ograniczeniem jest to, że rzeczywiste gry rzadko są grami o sumie zerowej (gdzie wypłaty graczy sumują się do zera) i często bierze w nich udział więcej niż dwóch uczestników. Analiza takich gier jest bardzo niedoskonała i właściwie składa się z serii rozwiązań cząstkowych. Bywa, że istnieje najlepsze rozwiązanie dla obu graczy, ale przed sformułowaniem teorii gier nie było jasne, jak wyznaczać je w sposób matematyczny. Dopiero po 1944 roku intuicyjne rozwiązania problemów tego typu nabrało formalnych znamion teorii, a szczególny postęp miał miejsce w latach siedemdziesiątych.

U podstaw teorii gier znajduje się kilka podstawowych pojęć, które przyjmowane są jako założenia i podaje się tylko ich definicje. W tej pracy przyjmuję taką samą strategię: będę się posługiwać zdefinio-

wanymi przez siebie terminami, niemniej jednak pominię wszelkie filozoficzne problemy dotyczące nazywanych przez nie fenomenów, gdyż, choć oczywiście zadanie takie mogłoby okazać się niezwykle interesujące, jest to przedsięwzięciem znacznie wykraczającym poza ramy krótkiego opracowania.

I. 2. WAŻNE POJĘCIA: GRA, STRATEGIA, RACJONALNOŚĆ I UŻYTECZNOŚĆ

Teoria gier może być wykorzystywana do analizowania wszelkich zachowań, w których agens dąży do maksymalizowania swoich zysków. Wykracza zatem poza ekonomię, a nawet filozofię. Bywa wykorzystywana w wojskowości, negocjacjach, analizie zachowań społecznych, antropologii, czyli wszędzie tam, gdzie występują sytuacje konfliktu bądź kooperacji.

I. 2a. GRA

Używając aparatury proponowanej przez twórców teorii gier należy pamiętać, że została ona stworzona, by umożliwić uproszczony opis skomplikowanych bezpośrednich interakcji pomiędzy poszczególnymi członkami społeczeństwa, i dlatego słowo „gra” powinno być traktowane jako metafora naukowa. Philip D. Straffin⁴ podaje cztery warunki umożliwiające zastosowanie teoriogrowej analizy:

- Można wskazać co najmniej dwóch graczy. Graczem może być człowiek, ale także firma, państwo, czy nawet gatunek w znaczeniu biologicznym.
- Każdy gracz ma do wyboru pewną liczbę możliwych *strategii*, określających sposób rozgrywania przez niego gry.
- *Wynik* gry jest determinowany przez kombinację strategii wybranych przez poszczególnych graczy.
- Każdemu możliwemu wynikowi gry odpowiada zestaw *wypłat* dla poszczególnych graczy, których wysokość można wyrazić liczbowo.

Gracze, dążąc do zmaksymalizowania swoich zysków, wybierają strategię i tym samym wpływają na przebieg i wynik gry, nigdy jednak nie jest to wynik niezależny od zachowań pozostałych graczy. Zależności między graczami mogą kształtować się pomiędzy czystym konfliktem, gdy zysk jednej ze stron oznacza automatyczną stratę drugiej,

⁴ P. D. Straffin, *Teoria gier*, Warszawa 2004.

a kooperacją, gdy wspólne działanie daje wynik dla wszystkich lepszy niż działanie „na własną rękę”.

Dla jaśniejszego wyłożenia mechanizmów teorii gier, powołałam się na kilka prostych przykładów⁵. Rozważmy dwuosobową grę o sumie zerowej, czyli taką, w której wypłaty graczy sumują się do zera, w której udział bierze Pani Kolumna i Pan Wiersz (w każdym przypadku pierwsza wartość oznacza wypłatę Pana Wiersza, druga – wypłatę Pani Kolumny).

	A	B
A	(2, -2)	(-3, 3)
B	(0, 0)	(2, -2)
C	(-5, 5)	(10, -10)

Gra 1

Pani Kolumna może w tej grze zdecydować się na granie jednej z dwóch strategii: A lub B, natomiast Pan Wiersz ma trzy możliwości: A, B lub C. Gra 1. nie ma punktu siodłowego.

Definicja 2⁶: Wynik gry nazywamy punktem siodłowym, jeśli jego wartość jest mniejsza lub równa każdej wartości w jego wierszu oraz większa lub równa każdej wartości w jego kolumnie.

Oczywiście, jeśli gra ma punkt siodłowy, czyli optymalne rozwiązanie, to racjonalni gracze powinni wybierać strategie go zawierające, jeśli zaś go nie ma, to jej wynik zależy od wyboru strategii mieszanej, która w jednoznaczny sposób wskaże, które z rozwiązań będzie najbardziej efektywne dla każdego z graczy. Gry o sumie zerowej są specjalnym przypadkiem – dużo częściej mamy do czynienia z grami, gdzie wypłaty nie sumują się do zera, czyli występuje połączenie konfliktu i kooperacji. Rozważmy Grę 2.

	A	B
A	(1, 1)	(-2, 2)
B	(2, -2)	(-5, -5)

Gra 2

⁵ Przykłady gier czerpię z podręcznika P. Straffina, który proponuje analizę gier rozwiązywanych przez jego studentów, co umożliwi mi pokazanie ciekawych statystycznych prawidłowości.

⁶ Wszystkie definicje podawane w tekście pochodzą również z podręcznika P. Straffina, chyba, że zaznaczono inaczej.

W tej grze gracze mają tylko po dwie strategie, jednak widać, że rozumowanie, co zrobi druga strona, doprowadzi do najgorszego możliwego wyniku $(-5, 5)$. Rozumowanie przebiegałoby bowiem mniej więcej w ten sposób: jeśli Pani Kolumna spodziewa się, że Pan Wiersz zagra A, to powinna zagrać B. Jeśli Pan Wiersz przypuszcza, że Pani Kolumna zagra A, to powinien zagrać B i w ten sposób obie strony tracą. Od razu widać, że najlepszym wynikiem dla obu graczy jest $(1, 1)$, i zapewne, gdyby mogli się porozumieć w tej kwestii, powinni się na nie zdecydować. Jest ono może nieco zachowawcze, niemniej jednak gwarantuje, że żaden z graczy nie poniesie strat i jednocześnie wyjdzie z gry z minimalnym zyskiem. Tego typu gry mogą, ale niestety nie muszą dopuszczać komunikacji, może w nich też pojawić się „mediator”, który wprowadzałby zewnętrzny punkt widzenia i proponował sprawiedliwe rozwiązanie w przypadku, gdy strony nie mogłyby dojść do porozumienia samodzielnie. Dodatkowo należy podkreślić, że w grach n -osobowych ($n > 2$) istnieje możliwość zawiązywania koalicji i dzielenia wygranej pomiędzy członków zwycięskiej koalicji. Gry tego typu, jako zbyt skomplikowane, nie będą jednak przedmiotem mojej analizy.

I. 2b. STRATEGIA

Rozważmy kolejną grę (dla uproszczenia w tabeli podane są tylko wypłaty Pana Wiersza, ale ponieważ jest to gra o sumie zerowej, bardzo łatwo można odczytać z niej wypłaty Pani Kolumny), która pozwoli na zdefiniowanie kryterium dominacji. Kiedy przedstawiona niżej gra była omawiana na zajęciach przez P. Straffina, okazało się, że niektóre strategie były wybierane znacząco rzadziej niż inne. Szczególnie strategia C Pani Kolumny nie budziła zainteresowania studentów, ale jeśli bliżej się jej przyjrzeć okaże się, że działa się tak w nieprzypadkowy sposób. Otóż wobec strategii C Pani Kolumny strategia B okazuje się bezwzględnie lepsza, czyli w każdej komórce strategii B wynik jest nie gorszy niż w przypadku strategii C. Możemy zatem powiedzieć, że strategia B dominuje strategię C.

Definicja 3: Strategię można określić jako pewien zamierzony ogólny plan, zgodnie z którym będzie toczyła się gra i który podpowiada graczowi w jaki sposób odpowiadać na zachowania pozostałych uczestników. Jeśli zilustrować grę w formie drzewka, to strategię można wyjaśnić jako program mówiący graczowi znajdującemu się aktualnie w węźle, czyli punkcie wyboru, którą gałąź wybrać.

Definicja ta komplikuje się, jeśli odróżnimy od siebie gry o pełnej informacji od gier o tzw. niepełnej informacji. Przykładem pierwszego typu gier są np. szachy, gdzie gracze znają wszystkie reguły a także w konkretnej partii znają wszystkie poprzednie ruchy, zarówno swoje jak i przeciwnika. Inaczej jest w przypadku przytoczonej wcześniej gry „z mostami”, w której mamy ukryty parametr, który znacząco wpływa na wybieraną strategię.

Definicja 4: Strategia S dominuje strategię T, jeżeli każdy wynik dawany przez S jest co najmniej równie korzystny, co odpowiedni wynik dawany przez T, a przynajmniej jeden wynik dawany przez S jest bardziej korzystny niż odpowiedni wynik dawany przez T.

	A	B	C	D
A	12	-1	1	0
B	5	1	7	-20
C	3	2	4	3
D	-16	0	0	16

Gra 3

Kryterium dominacji pozwala na wykluczenie strategii zdominowanych i utworzenie mniejszej „podgry.” Mówi ono, że racjonalny gracz nigdy nie wybiera strategii zdominowanej.

Ponieważ nie zawsze gracze dokonują wyboru strategii równocześnie, ciekawą komplikacją jest uwzględnienie tego, który z graczy wykonuje ruch w grze jako pierwszy. Otóż w grach o sumie zerowej rozpoczynanie rozgrywki jest zawsze niekorzystne, natomiast w pozostałych grach zależy to od dodatkowych czynników.

Dodatkowo możemy dokonać jeszcze rozróżnienia pomiędzy strategiami czystymi, w których gracz wybiera bez losowania jedną dowolną strategię oraz strategiami mieszаныmi, w których gracz wybiera losowo jedną spośród strategii i gra zgodnie z nią z określonym uprzednio prawdopodobieństwem. Zazwyczaj, z wyjątkiem aberracji językowych, nie stosujemy w komunikacji strategii mieszanych. Jest tak głównie dlatego, że pozostawianie rozstrzygnięcia w istotnych kwestiach losowi jest dla większości użytkowników języka nie do zaakceptowania. Dlatego nie będą one istotne w tej pracy.

I. 2c. RACJONALNOŚĆ

Założenie o racjonalności jest fundamentem teorii gier. Bez niego nie może spełniać ona swojej eksplikującej roli. Nie będę jednak zajmować się podawaniem uzasadnień, ani sytuowaniem racjonalności w biologicznym wyposażeniu człowieka. Chcę tylko zwrócić uwagę, że najprościej, choć w sposób bardzo kontrowersyjny, można zdefiniować ją na użytek tej pracy jako *linguistic teleology*, przy czym teleologię należy rozumieć tu w wyjątkowo słabym, lokalnym sensie, jako nakierowanie na rezultat danej czynności dyskursywnej, czyli porozumienie. Nie dotyczy ona tym samym czynności dyskursywnych w ogóle, ale jedynie dowolnej określonej [czynności], którą aktualnie badamy, lub w której uczestniczymy, przy czym w każdym z tych przypadków cel wyznaczany jest indywidualnie. W takim pojęciu zawarte już są *implicite* niezbędne dla poprawnego funkcjonowania teorii założenia dodatkowe, takie jak normatywność i użyteczność, które jednak zostaną zdefiniowane osobno. Sądzę, że takie użycie słowa „racjonalność” pozwala na podtrzymanie twierdzenia, że ludzie, jeśli idzie o podstawowe zdolności, a do nich należy zaliczyć umiejętność posługiwania się językiem, różnią się najwyżej stopniem biegłości. Niewykluczone oczywiście, że lepsze przyswojenie zasad języka pozwala nie tylko na sprawniejsze posługiwanie się nim, ale również na zdobywanie strategicznej przewagi już na początkowym etapie gry. Niemniej jednak, gdyby przedstawić sam język w postaci gry, sądzę, że byłby on grą o pełnej informacji. Dopiero toczące się za jego pomocą realne gry odgrywane przez użytkowników języka, mogą zawierać ukryty parametr, który komplikuje wybór strategii przez poszczególnych graczy.

Założenie o racjonalności można pokazać również, jako sumę kilku bardziej podstawowych założeń. Racjonalny gracz posiada cel. Ponadto może go osiągnąć oraz, dzięki posiadanym przez siebie danym, potrafi wybierać pomiędzy różnymi prowadzącymi do niego strategiami. Wcześniej pokazałam, czym ten wybór może być motywowany⁷. Racjonalność wymaga również, by, decydując się na granie jednej ze swoich strategii, uwzględniać możliwe wybory pozostałych uczestników. Innymi słowy, wymaga to świadomości brania udziału w grze. Jest to cecha, która łączy teorię gier z klasyczną ekonomią, gdzie wskazuje się na dążenie do maksymalizacji zysków, jako motywację działań ekonomicznych poszczególnych osób, nawet, jeśli zysk miałby być mierzony według subiektywnych kryteriów.

⁷ Por. poprzednia sekcja, strategie dominujące i zdominowane.

Okazuje się jednak, że na gruncie teorii gier rozróżnia się pomiędzy różnymi typami racjonalności. Podstawowe rozróżnienie mówi o racjonalności indywidualnej, czyli kryterium dominacji i racjonalności grupowej. W grach o sumie niezerowej odpowiednikami punktów siodłowych są tzw. równowagi Nasha, choć oczywiście, podobnie jak nie każda gra o sumie zerowej posiada punkt siodłowy, tak nie każda gra o sumie niezerowej ma równowagę. Rozważmy jednak następującą grę o sumie niezerowej:

	A	B
A	(3, 3)	(-1, 5)
B	(5, -1)	(0, 0)

Gra 4

Gra 4 ma jedną równowagę (B, B), co jest oczywiste tym bardziej, że dla obu graczy strategie A są zdominowane. Kiedy jednak przyjrzymy się diagramowi, okaże się, że nie jest to najlepsze rozwiązanie dla żadnego z nich. Wyraźnie widać, że lepiej byłoby grać (A, A). Około roku 1900 Vilfredo Pareto sformułował postulat, zgodnie z którym system ekonomiczny nie powinien być akceptowany, jeśli istnieje system dla wszystkich korzystniejszy. W obrębie teorii gier postulat ten przyjmowany jest jako *Kryterium Pareto* i łączący się z nim termin *paretooptymalność*.

Definicja 5: Wynik gry jest nieoptymalny w sensie Pareto, jeśli gra ma inny wynik, dający obu graczom wyższe wypłaty, lub jednemu z nich taką samą, a drugiemu wyższą. Wynik jest paretooptymalny, jeśli innego takiego wyniku nie ma.

Kryterium Pareto mówi, że tylko wynik optymalny w sensie Pareto może być akceptowany jako rozwiązanie gry.

Kryterium Pareto bywa też nazywane zasadą racjonalności grupowej. Jeśli jednak ponownie spojrzymy na diagram Gry 4, okaże się, że w tym przypadku zasada racjonalności grupowej wchodzi w oczywisty konflikt z bardziej podstawową zasadą racjonalności indywidualnej, którą jest kryterium dominacji.

Najsłynniejszą i najczęściej dyskutowaną współcześnie grą jest z pewnością tzw. Dylemat Więźnia⁸ (choć nie jest to typowa gra).

⁸ Gra ta została po raz pierwszy przedstawiona przez Melvina Dreshera i Merrilla Flooda i jest przykładem na to, że w grach o sumie niezerowej z jednym tylko punktem

	A	B
A	(0, 0)	(-2, 1)
B	(1, -2)	(-1, -1)

Gra 5

Graczami są dwaj aresztanci, którym policjanci przedstawiają pewien „układ”:

- jeśli jeden z nich przyzna się, drugi nie, to ten, który się przyznał, dostanie łagodny wyrok (wypłata 1), drugi zaś wysoki (wypłata -2);
- jeśli obydwoj się przyznają, dostaną łagodne wyroki (wypłata -1);
- z drugiej strony obaj aresztanci wiedzą, że nie ma żadnych dowodów przeciwko nim i jeśli żaden się nie przyzna, wyjdą na wolność (wypłata 0).⁹

Dylemat tkwi, podobnie jak w przypadku Gry 4, w tym, że równowaga wypada w punkcie (B, B), ponieważ każdemu z nich „opłaca się” przyznać, ale istnieje lepsze dla obu rozwiązanie mianowicie (A, A), w którym obaj wychodzą na wolność. Pomimo, że każdemu z więźniów opłaca się przyznać, to najlepszy wynik osiągną, jeśli obaj będą milczeć. Konflikt pomiędzy racjonalnością indywidualną (kryterium dominacji) a racjonalnością grupową (kryterium Pareto) nie daje się znieść, i zachowania „samolubnych” jednostek dbających tylko o własne interesy w tego typu sytuacjach prowadzi do wyniku niekorzystnego dla wszystkich, również dla nich samych. W tej konkretnej grze można jednak powiedzieć, że ustalenie wartości dla łagodnego wyroku (1), jako większego niż dla wyjścia na wolność (0) jest nieintuicyjne. Dlatego proponuję rozważyć tę grę w ogólnej postaci podanej przez Rapoportą i Chammaha, której ogólne założenia pozwalają na zagwarantowanie działania kryterium dominacji. Wartości podane w tabeli należy traktować jako zmienne, pomiędzy którymi zachodzą pewne zależności: $T > R > U > S$ oraz $R \geq (S + T)/2$. W każdej grze zgodnej z tym schematem mogą one być zastąpione przez dowolne

równowagi, równowaga ta może nie być paretooptymalna. Natomiast historyjkę o więźniach dopowiedział do niej Albert W. Tucker z Uniwersytetu Stanforda.

⁹ Grę tę przytaczam również dlatego, że jest ona równoważna Grze 4 i pozwala na jej zanalizowanie.

wartości. Schemat ten może być wykorzystywany do analizy wielu społecznych zachowań, w których w konflikt wchodzi obie wyżej wymienione zasady racjonalności.

	C	D
C	(R, R)	(S, T)
D	(T, S)	(U, U)

Gra 6

C – kooperacja (nie przyznawać się)

D – dezercja

R – nagroda za kooperację (*reward*)

S – wypłata frajera (*sucker*)

T – wypłata pokusy (*temptation*)

U – wypłata niekooperacyjna (*uncooperative*)

I. 2d. UŻYTECZNOŚĆ

Jednym z problemów, jakim musi sprostać teoria gier, jest problem kryterium przypisywania wynikom wartości. Jest jasne, że nie sposób podać uniwersalnego i obiektywnego kryterium tam, gdzie tak wiele zależy od subiektywnych uwarunkowań, a także od rodzaju gry i okoliczności, w których jest rozgrywana. By jednak w ogóle móc stosować teorię gier, konieczne trzeba założyć, że każdorazowo istnieje sposób przypisywania wartości wypłatom, a tym samym ustalenie preferencji poszczególnych graczy.

Do lat trzydziestych XX wieku użyteczność definiowano jako miarę subiektywnej satysfakcji. Jednak pod wpływem behawioryzmu, ekonomiści zrezygnowali z mówienia o nieobserwowalnych bytach i zaczęli szukać bardziej praktycznego kryterium dla użyteczności. W 1938 roku Paul Samuelson zdefiniował użyteczność (niestety w sposób cyrkularny) jako „działania, które podejmuje agent w celu zwiększenia użyteczności”. I choć wielu teoretyków wykorzystuje mimo to tę definicję, to jednak można również znaleźć inne, według mnie bardziej adekwatne. Użyteczność miałaby zagwarantować możliwość wystarczającego wyjaśnienia strategicznego rozumowania w odróżnieniu od strategicznego zachowania. Aby móc utrzymać formalny sposób rozumowania w obrębie teorii, konieczne jest posiadanie metody matematyzacji użyteczności, którą można by nazwać funkcją użyteczności. Jest ona odpowiedzialna za odtworzenie sposobu uporządkowania preferencji poszczególnych graczy. Wyobraźmy sobie prostą grę 3×2 , w której zamiast wartości liczbowych występują zmienne:

	A	B
A	u	v
B	w	x
C	y	z

Gra 7

Powiedzmy, że mamy do czynienia z grą o sumie zerowej – wtedy ustalenie preferencji jednego graczy pozwoli na automatyczne odtworzenie preferencji drugiego. Powiedzmy, że Pan Wiersz uporządkuje zmienne w następujący sposób: u, w, x, z, y, v , i że porządek ten jest przechodni. Dzięki temu można poszczególnym zmiennym przypisać konkretne wartości liczbowe:

	A	B
A	6	1
B	5	4
C	2	3

Gra 8

Gra 8 ma punkt siodłowy w postaci (B, B), i jak w każdej tego typu grze mamy do czynienia z użytecznością porządkową. Inaczej jest w przypadku gier, do których rozwiązania trzeba wykorzystać strategię mieszane. Wtedy konieczne jest ustalenie wartości w taki sposób, aby również proporcja między dwoma wypłatami była interpretowalna. Taką skalę nazywamy interwałową, a zamieszczone na niej użyteczności kardynałnymi. Do wyznaczenia ich wartości stosuje się schemat loterii, czyli pyta się gracza, czy wolałby, powiedzmy, na pewno uzyskać x , czy może raczej wziąć udział w loterii, w której z prawdopodobieństwem 0,5 otrzyma u oraz z prawdopodobieństwem 0,5 v . Jeśli gracz woli x od loterii, oznacza to, że średnia wartość v oraz u jest niższa niż wartość x . Postępując w ten sposób można uzupełnić oś wartości i znaleźć porządek użyteczności. Nie jest to jednak zadanie proste ani oczywiste, a gracze często nie potrafią jednoznacznie podać swoich preferencji w odniesieniu do różnych proponowanych im loterii. Można jednak, jak sądzę, założyć, że potrafiliby to zrobić, gdyby dać im na to nieco więcej czasu.

Dla Jeremiego Benthama użyteczność jest całościowym układem zysków i strat, które rozumiał jednak w dość pojemnych kategoriach

przyjemności i przykrości. Zatem zysk jest różnicą przyjemności i przykrości, czyli kosztów, które jednostka musi ponieść, dążąc do swojego celu. W odniesieniu do teorii gier nieskuteczne okazują się definicje w stylu dziewiętnastowiecznej ekonomii, mówiące o „maksymalizacji sumy użyteczności”, czy te bardziej współczesne, w których pojawia się hipotetyczny „człowiek racjonalny”¹⁰. W obrębie teorii gier użyteczność jest przypisywana pewnym działaniom tylko dlatego, że osoba, dla której mierzymy stopień użyteczności, jest skłonna częściej wybierać właśnie to działanie. Ponadto jeśli osoba wybiera jednak możliwość, którą określiliśmy jako mniej użyteczną, należy ponownie rozważyć, czy rzeczywiście we właściwy sposób ustaliliśmy dla niej ten porządek, a nie określać jej zachowanie jako irracjonalne. Trudności sprawiają zwłaszcza sytuacje, w których porównać trzeba wartości w jakimś sensie niewymierne i pochodzące z różnych porządków. Natomiast możliwość intersubiektywnego porównywania i sumowania użyteczności jest szkodliwym mitem, ponieważ nie można stopnia użyteczności oznaczyć inaczej, niż tylko na podstawie indywidualnych wyborów konkretnego gracza. Próby porównywania nie mogą być skuteczne, ponieważ nie można podać interpretacji dla otrzymanej informacji (jeśli można to jeszcze nazywać informacją). W grach o sumie zerowej możliwe jest zestawienie preferencji graczy tylko dlatego, że są one po prostu dokładnie przeciwstawne. Dzięki temu można podać najprostszą definicję wykorzystywanego tu pojęcia.

Definicja 6: Użyteczność jest najprostszą metodą systematyzowania informacji o preferencjach danego gracza.

Gracz przewidując możliwe kontynuacje gry, potrafi ocenić, które z nich są dla niego korzystniejsze i tym samym, które z nich preferuje a dane, w sposób arbitralny, uporządkować z uwzględnieniem przechodniości. Zapis: $u(A) > u(B)$ należy czytać jako: A jest preferowane nad B, ponadto, jeśli: $0 \leq p \leq 1$, to spełniony jest następujący warunek: $u[pA + (1-p)B] = pu(A) + (1-p)u(B)$, przy czym funkcja wyznaczająca użyteczności kardynalne jest liniowa. Użyteczność nie jest zatem żadną postulowaną wartością występującą obiektywnie, a jedynie nakła-

¹⁰ Co nie pozwala jednak całkowicie wykluczyć, że w niektórych przypadkach takie rozumienie może okazać się przydatne. Generalnie jednak teoria użyteczności sformułowana dla potrzeb teorii gier jest narzędziem wygodniejszym i o szerszym zakresie stosowalności.

¹¹ Gdzie p należy rozumieć jako wartość liczbową dla prawdopodobieństwa danego scenariusza.

danym na wybory graczy wygodnym schematem, czy, mówiąc inaczej, pewnym operacyjnym uproszczeniem pozwalającym mówić o osiągniętej przez nich satysfakcji.

II. GRY KOORDYNACYJNE I KONWENCJE

II. 1. CZYM JEST KOORDYNACJA?

Gry można roboczo podzielić na gry konfliktu bądź gry kooperacji, pomiędzy którymi jest całe spectrum gier pośrednich, niemniej jednak, z punktu widzenia Lewisa, teoria konwencji powinna oprzeć się o teorię gier czysto koordynacyjnych.

II. 1a. POWSTAWANIE I DEFINICJA PROBLEMÓW KOORDYNACYJNYCH

Definicja 7: Grą koordynacyjną nazywamy każdą i tylko taką grę, w której wypłaty graczy mają większą wartość, kiedy podejmują działania tego samego rodzaju, i w których zależność ta ma dla nich większe znaczenie niż same działania, które podejmują.

Przykładem zachowania podlegającemu takiej analizie jest ruch uliczny. Łatwo jednak zauważyć, że strategia „wszyscy jeżdżą prawą stroną” jest dokładnie tak samo punktem równowagi, jak „wszyscy jeżdżą lewą stroną”, a wyboru pomiędzy nimi nie można sprowadzić do analizy prawdopodobieństwa, z jakim należy je stosować. W takim przypadku możliwe strategie muszą podlegać racjonalizacji, czyli musi istnieć możliwość sprowadzenia ich do systemu przekonań, z którym punkt równowagi będzie spójny. Każdy gracz, podejmując swoją decyzję, bierze pod uwagę nie tyle to, gdzie jest *equilibrium*, ile raczej to, jaką decyzję podejmą pozostali. Dodatkowym argumentem za takim zachowaniem się może być to, że gry koordynacyjne bardzo często są wielokrotnie powtarzane, co znacząco wpływa na przyjmowaną strategię (podobnie jak u Hobbesa). Stronom po prostu opłaca się współdziałać, ponieważ preferują wyniki koordynacyjne.

Aby w ogóle mówić o grze, gracze muszą mieć wybór pomiędzy przynajmniej dwiema strategiami, a ich decyzje wpływają nie tylko na ich własne wypłaty, ale równocześnie na wypłaty pozostałych graczy. Kiedy więc w grze występuje *equilibrium*, a każda gra czystej koordynacji ma przynajmniej jedno, to żaden z graczy nie mógłby uzyskać lepszego dla siebie wyniku, chyba że co najmniej jeden z pozosta-

łych zagrałby inaczej. Osiągnięcie *equilibrium* jest więc możliwe tylko z pomocą zgodnych oczekiwań dowolnego rzędu dotyczących zarówno działań, jak i preferencji i racjonalności innych graczy.

II. 1b. ROZWIĄZYWANIE PROBLEMÓW KOORDYNACYJNYCH

Osiągnięcie sukcesu w grach koordynacyjnych nie jest, jak zresztą w żadnej z gier, zagwarantowane przez samo istnienie punktu równowagi. Z drugiej strony, *equilibrium* można osiągnąć przez przypadek, działając na ślepo, lub nie mogąc odgadnąć, jakie są preferencje pozostałych. Niemniej jednak wydaje się, że szanse na osiągnięcie punktu równowagi wzrastają wraz ze wzrostem wiedzy gracza o zgodności wspólnych oczekiwań. Jeśli gracz ma pewność, że pozostali gracze podejmą określone działania, to ma on wystarczający powód do tego, by również je podjąć. Może się jednak zdarzyć, że nie ma on, a nawet nie może mieć, pełnej informacji na temat ich preferencji – wtedy musi zdecydować, które z działań jest dla niego lepsze, ponieważ może istnieć inna strategia, którą, mimo że nie prowadzi do równowagi, uważa za korzystniejszą w przypadku, gdyby pozostali nie dążyli do *equilibrium*, ale również nie najgorszą, jeśli jednak się na nie zdecydują. W przypadku współzależnych decyzji kuszące może być zastanawianie się, co w tej sytuacji zrobi druga strona, ale niestety iterowanie takiego rozumowania nie doprowadzi do podjęcia żadnej konkluzywnej decyzji, ponieważ nieustannie będzie przybywać danych do analizy.

Przy takiej procedurze, zdaniem Lewisa,¹² korzystnie jest rozróżnić pomiędzy oczekiwaniami pierwszego rzędu i oczekiwaniami wyższych rzędów.

Definicja 8: Oczekiwaniem pierwszego rzędu nazywamy zwyczajne oczekiwanie dotyczące czegoś. Oczekiwaniem ($n+1$) rzędu nazywamy czyjeś przekonanie dotyczące przekonania n rzędu innej osoby.

Jeśli mój rozmówca spodziewa się, że po naciśnięciu włącznika zapali się żarówka, a ja znam jego przekonanie, to mam przekonanie pierwszego rzędu o jego przekonaniu oraz przekonanie drugiego rzędu w sprawie żarówki.

Do podjęcia decyzji nie musi jednak skłaniać gracza tylko jeden ciąg oczekiwań, przeważnie jest ich kilka, a im więcej możemy ich

¹² D. Lewis, *dz. cyt.*, s. 28.

znaleźć, tym nasza decyzja ma większy walor obiektywności i, co być może ważniejsze, pewniej prowadzi do celu. Najłatwiej podjąć taką decyzję, jeśli gra dopuszcza możliwość komunikacji pomiędzy graczami tak, że mogą oni uzgodnić między sobą, na którą z możliwych równowag decydują się. Dużo ciekawsze są jednak te gry, w których gracze muszą podjąć decyzję nie wiedząc jak zagrają pozostali. Lewis przywołuje badania T. C. Shellinga¹³, który badał zachowania graczy w grach koordynacyjnych bez możliwości komunikacji, przy założeniu, że rozumieją oni stawiane przed nimi zadanie w ten sam sposób, względnie dostali takie same instrukcje. Okazuje się, że osoby poddane eksperymentowi całkiem nieźle radziły sobie z rozwiązywaniem tych problemów, choć oczywiście ich skuteczność nie była równa tej osiągananej w przypadku gier już znanych, gdzie zawsze istnieje możliwość odwołania się do poprzednich rozgrywek stanowiących przesłankę w wyborze strategii. Załóżmy, że dwie osoby, które się nigdy wcześniej nie spotykały, chcą umówić się na spotkanie. Oczywiście mogą, na podstawie posiadanej już wiedzy, lub losowo, wybrać właściwe miejsce, niemniej jednak bezpieczniej byłoby, gdyby uzgodniły wcześniej tę kwestię między sobą. Inaczej w przypadku osób, które spotykają się regularnie w tych samych miejscach. Istnieje duże prawdopodobieństwo, że będą umawiać przy pomocy formuły: „Tam, gdzie zawsze, o godzinie...”, a z czasem nawet: „To może o godzinie...” i mimo to doskonale się rozumieją, ignorując nieskończoną ilość dopuszczalnych alternatywnych zachowań. Lewis twierdzi, że „siła perswazyjna” zachowań poprzedzających opiera się na *analogii*, z tym że, zaznacza, nie jest do końca jasne, na jakiej zasadzie budowane są sieci podobieństw. Ważne jednak, by pamiętać, że nie zawsze osoby, które ustanawiają „precedensowe” zachowanie są tymi samymi, które zgodnie z nim postępują.

II. 2. KONWENCJE JAKO OPIS PREFERENCJI GRACZY

W sytuacji, gdy poprzedzających gier jest więcej, gracz ma *warunkową preferencję* do dążenia do zgodności. Warunkową, ponieważ decyduje się na swój ruch tylko, jeśli pozostali gracze będą zachowywać się zgodnie z jego oczekiwaniami. Jest to jasne na mocy założeń teorii gier, gdzie racjonalność i użyteczność pozwalają oceniać preferencje biorących w grze osób. Ponieważ zaś uczymy się przez powtarzanie, seria wzmacniających zachowań poprzedzających ma większą „siłę przekonywania”, niż pojedyncze zdarzenie. Ponadto pojedynczy przy-

¹³ D. Lewis, dz. cyt., s. 13.

kład pozostaje niejednoznaczny, im jest ich więcej, tym większe szanse, że niejednoznaczności zostaną wyeliminowane. Jeśli liczba przykładów jest odpowiednio wysoka, poszczególni członkowie społeczności nie muszą bynajmniej znać dokładnie tych samych, ponieważ przy dużej ich liczbie łatwiej można wyeliminować niejasności komplikujące interpretację. Mechanizm wydaje się dość prosty: jeśli gracz zauważył, że do osiągnięcia koordynacji prowadzą konkretne zachowania, natomiast rzadko (lub wcale) spotykał się z tym, że ludzie zachowują się inaczej, będzie przypuszczał, że partnerzy w grze mają doświadczenie tego samego typu. Matematycznie można tę intuicję przedstawić w formie twierdzenia, udowodnionego przez Johna Nasha, że żaden punkt równowagi nie może znajdować się w obrębie strategii zdominowanej.¹⁴ Taka analiza stosuje się do przykładu z ruchem prawostronnym (lewostronnym), jako przykład można podać też „system flagowy” na amerykańskich skrzynkach pocztowych: jeśli flaga jest podniesiona listonosz wie, że właściciel zostawił tam list czekający na wysłanie, a skoro widzę, że listy moich sąsiadów zostają zabrane, gdy postępują w ten sposób, nie ma powodu, bym nie zrobiła tego samego.

Stała zgodność w określonych okolicznościach jest właśnie poszukiwanym punktem równowagi i dlatego każdy chce postępować tak jak inni.¹⁵ Zgodne postępowanie tworzy system oczekiwań, te zaś prowadzą do zgodnych zachowań, dlatego, zwłaszcza w obrębie języka, praktyka zwiększa szanse na skuteczność. Prowadzi to zdaniem Lewisa do najprostszej definicji konwencji:

Definicja 9¹⁶: Regularność R w zachowaniu członków populacji P , w pewnej rekurencyjnej sytuacji S nazywam konwencją wtedy i tylko wtedy, gdy każdorazowo w S spełnione są następujące warunki:

- (i) wszyscy podporządkowują się R ;
- (ii) wszyscy oczekują, że pozostali będą działać zgodnie z R ;
- (iii) każdy preferuje działanie zgodne z R , pod warunkiem, że inni też tak robią, ponieważ S jest traktowana jako problem koordynacyjny, zaś R stanowi koordynacyjne equilibrium dla S .

¹⁴ J. Nash, *Non-Cooperative Games*, „The Annals of Mathematics”, vol. 54, nr 2.

¹⁵ „Tak jak inni” nie oznacza działania dokładnie takiego samego, ale jedynie działanie zgodne z określonym schematem.

¹⁶ D. Lewis, *dz. cyt.*, s. 42.

Później Lewis osłabia nieco swoją definicję mówiąc, że wystarczy, by *prawie* wszyscy postępowali zgodnie z konwencją, ale, jak wykazała Ruth Millikan, nie jest konieczne, aby wszyscy członkowie społeczności działali w ten określony, regularny sposób. Problemem nie jest ustalenie jak dużą grupę muszą stanowić osoby znające konwencję, ale znalezienie dla niej warunku minimalnego. Wspominałam już w Części I, że niejednokrotnie wystarczy procentowo niewielka grupa członków, by mówić o konwencji. Przykładem może być odmiana czasownika „iść” dla rodzaju męskiego w języku polskim, pomimo że większość użytkowników odmienia niepoprawnie jako „poszłem”, konwencją w obrębie języka jest „poszedłem”. Należy jednak poczynić jedno zastrzeżenie: aby działać zgodnie z konwencją, należy świadomie dążyć do *equilibrium*. Można oczywiście wykonać określoną czynność przez przypadek, ale wówczas nie można mówić o działaniu konwencjonalnym, ponieważ pogwałcone zostaje kryterium racjonalności. Nie jest nim także działanie, które ma znamiona działania konwencjonalnego, ale wykonująca je osoba jest tak przywiązana do niego tak, że nawet, jeśli pozostali członkowie społeczności zmieniają swoje zachowanie, ona nadal będzie zachowywać się w ten sam sposób. Podstawą dla działania zgodnego z konwencją jest wspólna wiedza i wspólne doświadczenia. Regularność *R* może występować w pewnej społeczności, zanim zostanie w jakikolwiek sposób sformułowana, niemniej jednak, dopiero po uświadomieniu sobie tej prawidłowości można mówić o konwencji. Znajomość brzmienia konwencji nie jest jednak warunkiem koniecznym do tego, by zachowanie można było określić jako zgodne z nią. W procesie akwizycji języka użytkownicy skutecznie nabywają umiejętności posługiwania się językiem w sposób, który może im zapewnić porozumienie, a więc *equilibrium*, ale bynajmniej nie zawsze potrafią wysłowić reguły, które „moderują” użycie języka.

Nie jest bynajmniej tak, że odkrywając konwencje dostosowujemy do nich własne preferencje. Mechanizm wydaje się być zupełnie inny. Z pewnych, nie zawsze dających się jednoznacznie opisać, względów niektóre zachowania są wybierane częściej niż inne, dlatego możemy przypisać im wyższą użyteczność. Dzięki temu zaś uznajemy je za racjonalne, czyli lepiej służące, w zależności od sytuacji, jednostce bądź grupie. Stąd niedaleko już do uświadomienia sobie regularności i próby jej teoretycznego opisu jako konwencji. Opis ten może być skuteczny tylko o tyle, o ile w jego centrum znajdują się działania, a nie przypisywane im znaczenia. Niewątpliwie uczestnicy gry chcą osiągnąć jakiś cel – w omawianym przypadku jest nim koordynacyjny punkt równo-

wagi – niemniej jednak motywacje, które nimi kierują nie są i nie mogą być poddane analizie w ramach tej teorii. Aby opis był skuteczny, konieczna jest tylko świadomość, że gracze mają preferencje oraz, że potrafią uszeregować nadawane im użyteczności kardynalne, ale nie muszą tłumaczyć już dlaczego jedno z rozwiązań cenią bardziej od innych.

II. 3. KONWENCJE KOMUNIKACYJNE

II. 3a. KONWENCJE W SYSTEMACH ZNAKÓW

Zanim przejdę do omówienia zasadniczego sposobu stosowania konwencji, chcę zauważyć, że komunikacja nie odbywa się tylko z pomocą języka, bardzo wiele informacji przekazywanych jest w nieleksykalnych systemach znaków. Może to być sygnalizacja świetlna czy znaki drogowe. Wszystko jednak wskazuje na to, że, aby nadawać sygnały konieczne jest zapewnienie minimalnych warunków odbioru: każdorazowo warunki te mogą być inne, w zależności od rodzaju wykorzystywanych sygnałów, ale jeśli nie zostaną one spełnione nie można w ogóle mówić o nadawaniu komunikatu. Najczęściej zakłada się ich spełnienie, choć może być tak, że dopiero ustanawia się je przy pomocy metasystemu wyjaśniającego zasady przekazu, i traktuje jako elementy „wspólnej wiedzy”. Od strony formalnej można przedstawić to w następujący sposób: w pewnej sytuacji S , osoba, która chce przekazać komunikat N pojedynczemu odbiorcy (lub ich grupie) O , ma możliwość wyboru pomiędzy różnymi alternatywnymi sposobami przekazania informacji s_1, \dots, s_n , które będą miały równoważną treść, a odbiorcy potrafią określić, którą z czynności wykonał. Z drugiej strony, odbiorcy mają do wyboru szereg możliwych odpowiedzi r_1, \dots, r_m i zależnie od preferencji wybiorą jedną z nich. Lewis zakłada, że cała grupa odbiorców musi preferować tę samą odpowiedź, co, jak sądzę, jest dalece zbyt restrykcyjnym warunkiem: aby skutecznie nadać komunikat, wcale nie wszyscy muszą go zrozumieć, natomiast konieczne jest założenie, że w grupie adresatów znajdują się osoby, które potrafią zarówno zrozumieć komunikat jak i odpowiedzieć na niego. Zatem:

Definicja 10: $\prod s_j \in \{s_1, \dots, s_n\} \sum r_i \in \{r_1, \dots, r_m\}$ takie, że $\langle s_j; r_i \rangle$ należy traktować jak parę strategii dających żadaną równowagę w grze¹⁷.

¹⁷ Moja definicja stanowi modyfikację definicji podanej przez D. Lewisa, dz. cyt., s. 131. Dodatkowym warunkiem jest założenie, że zarówno nadawca N , jak i odbiorcy O są członkami społeczności, która zna dany kod.

Ponieważ jednak gry koordynacyjne mogą mieć więcej niż jedno equilibrium, możliwe jest wiele strategii, które, pomimo różnic formalnych, będą prowadzić do porozumienia. Wieloosobowa gra tego typu redukuje się do szeregu gier dwuosobowych, dlatego też nie wszyscy odbiorcy muszą odpowiedzieć tak samo.

II. 3b. KONWENCJONALNE GRY JĘZYKOWE

Aby nabyć zdolności posługiwania się językiem, trzeba najpierw odkryć regularności w jego obrębie, nie jest możliwe uczenie się języka, jeśli każdą kolejną wypowiedź traktuje się jako oderwaną od systemu. Uświadomienie sobie tej regularności jest pierwszym stopniem inicjacji w społeczności osób posługujących się danym językiem naturalnym. Nie sądzę jednak, aby konieczne było spełnienie nakładanego na członków społeczeństwa przez Lewisa warunku, mówiącego, że muszą oni wiedzieć, czym te konwencje są. Sądzę, że właśnie teoria gier umożliwia wycofanie się z tak restrykcyjnych poglądów. Konwencje, podobnie jak użyteczność są narzędziami, które ułatwiają opis regularności pewnego typu, ale przecież aby posługiwać się narzędziem nie trzeba znać jego teorii. Zatem, kiedy Lewis mówi, że użytkownicy wykazują się wiedzą o konwencjach, podejrzewam, że mówi więcej, niżby należało. Oczywiście użytkownicy języka wiedzą nie tyle, co powiedzą, ile co *mogą powiedzieć*, żeby nadal pozostać w obrębie społeczności użytkowników języka. Zagwarantowany jest tylko schemat użycia języka, ale nie jego konkretna realizacja.¹⁸ Osoby, których językiem ojczystym jest język polski, są, dopóki fakty temu nie zaprzeczają, rozpoznawani przez pozostałych użytkowników tego języka jako posługujący się tym samym zespołem konwencji. Aby opisać komunikację językową w obrębie danego języka J w sposób formalny wystarczy przekształcić nieco zbyt już ogólną definicję 10:

Definicja 10³: $\prod s_j \in \{s_1, \dots, s_n\} \sum r_i \in \{r_1, \dots, r_m\}$ (gdzie $\{s_1, \dots, s_n\}$ oraz $\{r_1, \dots, r_m\}$ są podzbiorami J) takie, że $\langle s_j; r_i \rangle$ należy traktować jako parę strategii dających żadaną równowagę w grze¹⁹.

¹⁸ D. Lewis, dz. cyt, s. 62.

¹⁹ Mogą zdarzyć się gry, w których jeden z graczy odpowiada inaczej niż werbalnie, np. poprzez działanie, przedstawiona definicja może być traktowana jako schemat „negocjacyjny”, gdzie negocjacje są szeroko rozumiane jako wszelkie formy komunikacji odbywającej się przy wyłącznym użyciu języka jako środka.

Definicja 11: Język traktuję jako zinterpretowany zbiór symboli wraz ze skończonym zbiorem reguł gramatycznych rządzących ich łączeniem w większe zespoły treściowe oraz użyciem.

Podobną definicję przyjmuje Lewis, przy czym zaznacza on, że nie może być dwóch języków o tej samej interpretacji, niemniej jednak w dwóch językach mogą występować równobrzmiące wyrażenia²⁰. Język może zawierać wyrażenia różnego typu, łącznie z modalnymi oraz indeksykalnymi i, aby spełniał swoje cele, musi być traktowany jako nieskończony. Mając zadaną interpretację języka, jego użytkownicy wykorzystują go w pewien określony sposób, i nawet, jeśli występują niejednoznaczności, to są one związane z mnogością sposobów wyrażania tych samych treści za pomocą bogatych środków zawartych w systemie. Konwencje nie mogą być ani zbyt szerokie (wtedy trudno wyselekcjonować zbyt liczny zbiór ich zastosowań), ani zbyt wąskie (ponieważ mała liczba przykładów również komplikuje akwizycję). Dziecko uczy się języka przez dostrzeganie cech wspólnych w ograniczonej próbie językowej, ale im większe ma doświadczenie, tym mniej popełnia błędów i stopniowo staje się pełnoprawnym członkiem społeczności użytkowników danego języka naturalnego. Z tego też powodu John H. Nachbar²¹ wyróżnia strategie konwencjonalne:

Definicja 12: Skojarzony z każdym z graczy podzbiór powtarzalnych w grze strategii nazywamy strategią konwencjonalną.

Gracze uczą się osiągać equilibrium nie tylko dzięki powtarzalności własnych decyzji, ale przede wszystkim dzięki pozostałym graczom, którzy zachowują się sposób regularny. Pozwala to na przewidywanie kontynuacji gry i tym samym optymalny dobór własnej strategii. Strategia nie może jednak być optymalna, jeśli gracz może nauczyć się przewidywać, że w obrębie kontynuacji danej gry, da ono rozwiązanie gorsze od optymalnego. Jeśli jednak nawet wielokrotne powtarzanie gry nie prowadzi do sformułowania konwencjonalnej strategii, gracze muszą działać innowacyjnie wykorzystując inne znane im narzędzia językowe.

W obrębie języka odbywa się wiele gier, najpowszechniejszym i dość prostym schematem jest zapewne gra w pytania i odpowiedzi. Aby taka gra mogła w ogóle zostać przeprowadzona, musi zostać speł-

²⁰ D. Lewis, dz. cyt., s. 162.

²¹ J. H. Nachbar, *Prediction, Optimization and Learning in Repeated Games*, „Econometrica”, vol. 65, nr 2.

niony szereg podstawowych warunków, które zostały wyszczególnione w poprzednich sekcjach tej pracy: zakładamy, że uczestnicy gry rozumieją ją w ten sam sposób, posługują się tym samym językiem, spełniają kryteria racjonalności oraz potrafią przypisać wartości swoim ewentualnym wypłatom. Punkt, w którym osiągają equilibrium jest w każdej rozgrywce inny, niemniej jednak jego osiągnięcie jest współzależne od decyzji, jakie każdy z nich podejmie. Strategie wybierane przez graczy można traktować jako entki uporządkowane (w zależności od tego, ilu osobowa jest gra), wskazujące na ich wypłaty.

Zachowania językowe, podobnie zresztą jak wszystkie zachowania, które można określić jako konwencjonalne, generują użyteczności jako wtórne elementy opisu, pomimo że nie mamy możliwości odtworzenia procesu kształtowania się preferencji członków społeczności, która wytworzyła daną konwencję. Istotne jest jednak, aby dostrzec, że każdorazowo za próbą użycia konwencji stoi zasada kooperacji i dążenie do osiągnięcia equilibrium. Jeśli tak zrozumiemy konwencję, oczywiste stanie się, że graczom generalnie nie opłacają się zachowania niezgodne z nimi, zwłaszcza na poziomie językowym. W przypadku bowiem innych gier może być tak, że poszczególni gracze mogą osiągnąć korzystniejsze wyniki, grając zgodnie z zasadą racjonalności indywidualnej. Jednak w obrębie języka sytuacje, w których użytkownicy nie chcą zostać zrozumiani (bo to jest równowaga) należy traktować jako aberracje, a wręcz pogwałcenie podstawowego celu języka, które jest wtórne względem jego podstawowej funkcji.

WNIOSKI

Podtrzymanie konwencjonalizmu dzięki teorii gier możliwe jest przede wszystkim dlatego, że spełnia ona dwa warunki: jest zarówno opisem, jak i uzasadnieniem. Z jednej strony pozwala na analizę konkretnych zachowań językowych, w konkretnych sytuacjach, z drugiej daje podstawy dla generalizacji poszczególnych wyników. Te dwa podejścia nie tylko się uzupełniają, ale jednocześnie wzajemnie usprawiedliwiają. Żadna teoria, która spełnia tylko jeden z tych warunków, nie może podtrzymać swoich roszczeń do kompletności. Dopiero w obrębie takiego pełnego systemu możliwe jest wyjaśnienie normatywności, która (podobnie jak użyteczność) jest pewnym uproszczonym schematem. Możemy, analizując problemy koordynacyjne, przewidywać zachowania graczy dzięki wykorzystaniu zasady racjonalności oraz kryterium użyteczności: jeśli dobrze przypiszemy użyteczności kardy-

nalne, wyznaczenie „normy” nie powinno stanowić trudności. Normą w tym wypadku jest bowiem dążenie do punktu równowagi, pozostałe zachowania należy traktować jako wypaczenie systemu, które jednak może być w jego ramach zbadane.

Zaletą analizy teoriogrowej jest także to, że pozwala na odparcie zarzutu arbitralności. Jest tak nie tylko dlatego, że wybór strategii w grze jest umotywowany psychologicznie, ale przede wszystkim dlatego, że każda nietrywialna gra ma pewną określoną liczbę strategii, spośród których się wybiera. Zatem gracz, jeśli chce grać w jakąś konkretną grę, nie może zachować się zupełnie dowolnie. Konieczne jest wybieranie strategii z określonego zbioru tak, by inni gracze (użytkownicy języka) mogli rozpoznać w nim partnera. Arbitralność została zatem zupełnie wyeliminowana, ponieważ wystarczy sprawdzić, czy ciągle jeszcze toczy się dana gra, czy już zostaliśmy wyrzuceni w obręb innej.

Mankamentem analizy teoriogrowej jest ukryte odwołanie do holizmu, które pojawia się, ponieważ trudno rozstrzygnąć, jak niewielka musi być zmiana, by już nie miała wpływu na wybór strategii. Szczególnie wyraźnie widać to tam, gdzie konieczne staje się odwołanie do wspólnej, czy powszechnej wiedzy: po pierwsze, jak sądzę, nie można podać pełnego jej pochodzenia i sposobu nabywania, konieczne jest przyjęcie jej jako danej. Po drugie, wątpliwe są metody umożliwiające łączenie mniejszych jednostek w sieci zależności. Na obecnym etapie trzeba przyznać, że pomimo pewnych ogólnych kryteriów, w dużej mierze decydują o nich czynniki subiektywne. Możliwe rozwiązanie tego problemu podaje John Nash: w abstrakcyjnych grach, aby osiągnąć określone rozwiązanie, potrzeba tylko minimalnej informacji. Być może jednak to ograniczenie nie stosuje się do realnych gier²². Nie jest też bynajmniej tak, jak twierdzi Ariel Rubinstein²³, że analiza w ramach teorii gier jest kusząca dlatego, że umożliwia badanie funkcjonowania umysłu. Sądzę, że jest właśnie odwrotnie: ujęcie to pozwala nam uniknąć wszelkich trudności, na jakie napotyka filozofia umysłu, a jednocześnie nadal zachowuje odróżnienie człowieka i maszyny.

Zarzutem stawianym konwencjonalizmowi może być również to, że nie wszystkie elementy języka są konwencjonalne. W sposób oczywisty niektóre spośród nich należą do biologicznego wyposażenia człowieka, inne warunkowane są przez wpływ środowiska. Z drugiej

²² J. Nash, *Two-Person Cooperative Games*, „Econometrica”, vol. 21, Nr 1.

²³ A. Rubinstein, *Comments on the Interpretation of Game Theory*, „Econometrica”, vol. 59, Nr 4.

jednak strony użycie języka jest wyznaczone przez pewne schematy przystosowawcze, do których należy zaliczyć koordynacyjne podejście do komunikacji. Konwencja jest regularnością zachowania (użycia), dlatego nie podejmuję tu analizy semantycznej. Świadomie nie zajmuję się również warunkami prawdziwości, pomimo tego, że Lewis poświęcił osobny rozdział na analizę semantyki możliwych światów. Oczywiście można powiedzieć, że gry w obrębie języka mogą dotyczyć kwestii znaczenia, ale sądzę, że wystarczająco często jest tak, że samo działanie wystarczy do tego, by zdefiniować znaczenie użytych słów. Myślę, że w wielu przypadkach możliwe jest zredukowanie semantyki do pragmatyki, a czy jest ono możliwe generalnie pozostaje otwartym programem badawczym.

THE CONVENTIONAL APPROACH TO STRATEGY IN SOLVING PROBLEMS OF COORDINATION

Summary

This article shows how communication through language can be expressed in terms of game theory. The general idea and the main line of argumentation is based on David Lewis' book *Convention* (Lewis 1969) and more recent works on game theory that develop the concepts of cooperation and equilibrium. The term *agreement*, in the sense given to it by game theory, is used to show how to justify the thesis of the conventional nature of language. In the first part of the article some general notions of game theory are presented. Game theory is supposed to examine strategies chosen by rational agents in order to obtain preferred outcomes and is thus a good tool for describing and predicting human behaviour. It applies to situations when subjects operate reasonably, taking into account the behaviour of other rational subjects, and thus can also be used in situations of language communication. Sustaining conventionalism with game theory is possible primarily because it satisfies two basic conditions: it is both a description and a justification. Only within a complex system is it possible to explain the normative character of language, which, like utility, is only a simplified scheme. Through the analysis of coordination problems and the use of the principles of rationality and utility, it is possible to anticipate the behaviour of agents. In this context normativeness should be regarded as striving for balance, and even if another way of understanding is a deviation of the system, it still can be explored and explained within it.

Magdalena Adamus