THE LOGIC OF SOCIAL RESEARCH

UNIVERSITY OF BIAŁYSTOK
Białystok 2004

THE LOGIC OF SOCIAL RESEARCH

Series: STUDIES IN LOGIC, GRAMMAR AND RHETORIC 7(20)

THE LOGIC OF SOCIAL RESEARCH

UNIVERSITY OF BIAŁYSTOK
Białystok 2004

Series: STUDIES IN LOGIC, GRAMMAR AND RHETORIC
edited by:
The Chair of Logic, Informatics
and Philosophy of Science (Faculty of Mathematics and Physics)
Department of Semiotics (Faculty of Law)
University of Białystok
e-mail: logic@uwb.edu.pl

Editor for the volume: **Halina Święczkowska**
Editorial Assistants: Katarzyna Doliwa, Dariusz Surowik

Refereed: Witold Marciszewski, Kazimierz Trzęsicki

Cover design: Krzysztof Michałowski
Type-setting: Stanisław Żukowski

This volume was supported by Komitet Badań Naukowych

ISBN 83-7431-015-4
ISSN 0860-150X

WYDAWNICTWO UNIWERSYTETU W BIAŁYMSTOKU
15-097 Białystok, ul. Marii Skłodowskiej-Curie 14, tel. (085) 7457059
http://wydawnictwo.uwb.edu.pl, e-mail: ac-dw@uwb.edu.pl
Nakład 300 egz.

Druk i oprawa: Podlaska Spółdzielnia Produkcyjno-Handlowo-Usługowa
Białystok, ul. 27 Lipca 40/3, tel./fax 6754802
http://www.podlaska.com.pl

CONTENTS

Witold Marciszewski
Uniwersytet w Białymstoku

WSPOMNIENIA I PROJEKTY JUBILEUSZOWE Z OKAZJI DWUDZIESTEGO TOMU ROCZNIKA *STUDIES IN LOGIC, GRAMMAR AND RHETORIC*

Na prawach jubileuszowego wyjątku tekst ten ukazuje się po polsku, choć angielski nie jest dla polskich czytelników barierą. W języku jednak macierzystym pióro bardziej giętko „odda to wszystko, co pomyśli głowa". Głowa zaś chciałaby podzielić się racjami, z których to pismo się zrodziło, jak i wizją jego nowej fazy.

Tytułem swym pismo nawiązuje do średniowiecznego Trivium, niech więc wolno mi będzie przedstawić się na wstępie scholastycznym przydomkiem *Inceptor* czyli początkodawca (w pełnym brzmieniu *Inceptor Venerabilis*). Niecodzienna myśl utworzenia w Białymstoku pod tym tytułem pisma w językach kongresowych, głównie angielskim (choć *genius loci* podpowiadałby raczej jakiś idiom kresowy), pojawiła się w roku 1979. Data 1979 jest znacząca. Polityka bowiem trzymała wtedy życie umysłowe w tak mocnym uścisku, że prościej było założyć pismo w drugim czyli nielegalnym obiegu, niż uzyskać u władz legalizację nowego projektu piśmienniczego.

Na początek wypada zdać sprawę, skąd ten pomysł przywołania w tytule średniowiecznego Trivium. Jest to zarazem sposobność, żeby oddać sprawiedliwość dwom znaczącym postaciom w życiu akademickim tamtego czasu, z którymi miałem przywilej współpracować. Jedną z nich jest Jerzy Pelc, pionier semiotyki na gruncie polskim i wielce znacząca pozycja w międzynarodowym ruchu semiotycznym. To za jego sprawą stałem się uczestnikiem ruchu semiotycznego w Polsce i szerzej na polu zarówno organizacyjnym jak piśmienniczym, czego refleksem był pomysł nawiązania do dawnych dyscyplin semiotycznych (nazywanych *scientiae sermocinales* – logiki, gramatyki i retoryki.

Skąd jednak ów sztafaż średniowieczny zamiast wprowadzenia do tytułu nowoczesnego pojęcia semiotyki? Jeden motyw był negatywny, drugi pozytywny.

ISBN 83-7431-015-4 ISSN 0860-150X

Negatywny to ten, że mimo zaangażowania w tematykę semiotyczną miałem (i mam do dziś) opory przeciw wyodrębnianiu semiotyki jako osobnej nauki. Opory są natury ogólnej i dziś rozciągam je także na modną kognitywistykę. Rzecz w tym, że mnożenie nowych dyscyplin, ponoć w imię interdyscyplinarności, bywa działaniem fikcyjnym oraz rodzącym ryzyko obniżenia poziomu badań. Może to być o tyle fikcja, że realna współpraca między dyscyplinami bierze się z istniejącej potrzeby oraz umiejętności znajdowania wspólnego języka. A jeżeli jest potrzeba i umiejętności, to dotychczasowa klasyfikacja nauk i związane z nią struktury organizacyjne nauki często wystarczają do celów kooperacyjnych. Nie na darmo uniwersytet nazywa się tak od słowa *universitas*, a podobnie uniwersalne są PAN, PAU czy TNW. Sąsiedztwo w takiej jednej strukturze przeważającej większości nauk, choćby w ramach senatu, stwarza dobre warunki kontaktu. Trudno natomiast zaobserwować, żeby dzięki powołaniu jakiegoś instytutu semiotyki wzrosły kompetencje matematyczne zatrudnionych w nim lingwistów czy lingwistyczne socjologów; a tylko tego rodzaju wzrost służyłby współpracy między naukami (jeśli uczeni koledzy nie mają rozmawiać jak Chińczyk z Eskimosem). Innym argumentem przeciw pączkowaniu dyscyplin interdyscyplinarnych jest to, że te z przedrostkiem „inter" nie mają wypracowanego warsztatu badawczego ani przestrzeganych przez środowisko rygorów badawczych. Powiedzmy, w tzw. kognitywistyce przeciętny lingwista będzie siłą rzeczy najlepszym lingwistą wśród tworzących to nowe środowisko przybyszy z psychologii, pedagogiki, neurofizjologii itd. Podobnie, słaby informatyk czy neurofizjolog będzie tam brylował jako najwyższy autorytet informatyczny czy biologiczny, podczas gdy w ojczystych środowiskach każdy z nich jest kontrolowany wedle sprawdzonych od dawna standardów i porównywany z najlepszymi reprezentantami danej dziedziny. Słowem, taka instytucjonalna interdyscyplinarność może i pomaga karierom akademickim, ale gorzej służy nauce.

Przy tej sceptycznej postawie wobec instytucji z przydomkiem „semiotyczna", nie brałem pod uwagę użycia go w nazwie zamierzonego pisma. Ale, jako się rzekło, tematyka podpadająca pod ten termin jest realna i doniosła, będąc rozdzielona między kilka nauk o wielkiej tradycji, a w tym logikę, lingwistykę i naukę o komunikacji społecznej; jako reprezentacja (*pars pro toto*) tej drugiej dobrze się nadaje gramatyka, trzeciej zaś – okryta klasyczną patyną retoryka.

A co do motywacji pozytywnej, to wiąże się ona z inną postacią akademicką, której zawdzięczam szczególnie wiele inspiracji. W filozofię średniowieczną wprowadzał mnie Stefan Swieżawski; po latach pomogło mi to wydatnie w dotarciu do średniowiecznych prekursorów formalizmu w logice

(z czego wzięła się duża część książki, napisanej wspólnie z Romanem Murawskim, *Mechanization of Reasoning in a Historical Perspective*, Amsterdam 1995). Od Profesora Swieżawskiego promieniowała wizja średniowiecza jako epoki pulsującej wiarą w szanse poznania świata. To, że wobec ogromu zadań poznawczych uciekano się pod skrzydła autorytetów nie umniejszało przekonania, że rozum ludzki musi się sam potężnie wysilić. A do tej pracy przygotowują go trzy mądre korepetytorki: logika, gramatyka i retoryka. Po nich miało przyjść studium matematyki i nauk o przyrodzie, a potem studium teologii. Odziedziczywszy taką wizję, skromny logik, mający przy tym zainteresowania językowe, społeczne i historyczne, oraz pragnący być społecznie użytecznym (o taką bowiem bym się ubiegał intelektualną metrykę) czuł się w owym Trivium nauk jak u siebie w domu. Pragnął też przekazać to nastawienie współpracownikom, młodszym kolegom ze środowiska logiczno-informatycznego powstającego w uczelni i okolicy. Zaproponował więc im wspólne wydawanie rocznika: *Studies in Logic, Grammar and Rhetoric*.

*

* *

Od tomu 14 (rok 1997) pismo przeszło na nową formułę edytorską, łącząc to ze zmianą formy graficznej i uzyskaniem podwójnego statusu, periodyku i serii książkowej. Formalnie przejawia się w posiadaniu numerów ISBN i ISSN, a merytorycznie w tym, że każdy tom ma swój odrębny temat i wyrażający go tytuł, o czym dokładniej opowiadam niżej. Utrzymywanie tej formuły nie jest jakimś dogmatem, ma ona swoje za i przeciw; jest to kolejny temat do refleksji jubileuszowej.

Od roku 2001 pismo ma Radę Naukową (Editorial Advisory Board), w której znaleźli się czołowi przedstawiciele z głównych środowisk logiczno-filozoficznych w Polsce. Cieszy się też przynależeniem do grupy A czasopism rekomendowanych przez KBN. Jest to strategicznie dobry punkt wyjścia do ekspansji, której projekt zarysuję dalej.

Na czym polega nowa formuła edytorska, ta od roku 1997? Otóż każdy tom ma od tamtego roku określoną tematykę, wyrażoną w jego tytule. Zaczęło się to trochę przypadkowo, bo w roku 1996 poczuliśmy się powołani do tego, żeby uczcić 350-lecie urodzin Leibniza. Dzięki członkostwu w Leibniz Gesellschaft miałem sposobność uczestniczenia, także w formie wkładu pisarskiego, w obchodach tej rocznicy w Niemczech, były więc już gotowe pewne tematy i przemyślenia, zaś wieloletni Redaktor Naczelny *Studies* miał wcielenie w osobie Haliny Święczkowskiej, która właśnie kończyła swą rozprawę habilitacyjną *Harmonia Linguarum. Język i jego funkcje w filozofii Leibniza*. I tak w następnym roku ukazał się tom pt. *On Leibniz's Phi-*

losophical Legacy in the 350th Anniversary of His Birth, w którym do leibnizjańskiego wątku dołączyli się jeszcze inni autorzy z kraju i z zagranicy.

Następne tomy okazały się także mieć proweniencje rocznicowe, ponieważ tak się zbiegło, że wydająca *Studies* Katedra Logiki, Informatyki i Filozofii Nauki UwB (Uniwersytetu w Białymstoku) zaczęła od 1997 organizować co rok ogólnopolskie *Warsztaty Logiki, Informatyki i Filozofii Nauki*, każdy poświęcony jakiejś ważnej rocznicy historycznej w dziejach logicznej problematyki rozstrzygalności i złożoności obliczeniowej. Na przykład: stulecie urodzin Emila Posta i takież Johna von Neumanna, stulecie programu Hilberta, pięćdziesięciolecie Turinga idei sztucznej inteligencji, dziesięciolecie śmierci Friedricha Hayeka (zasłużonego dla teorii wolnego rynku jako systemu obliczeniowego). Jako wykładowcy zapraszani byli najlepsi polscy specjaliści od danego zagadnienia. Byłoby więc marnotrawstwem nie wykorzystać ich wkładu jako publikacji w wersji angielskiej, przygotowanej specjalnie dla *Studies*. Tak ukształtowała się, jakby na zasadzie samorzutnego porządku (Hayeka *spontaneous order*), pewna koncepcja programowa, którą może się, choć nie musi, kontynuować i doskonalić. Jest to kolejny temat do jubileuszowej dyskusji.

Na prawach głosu w tej dyskusji, wypowiadam się na rzecz kontynuacji dotychczasowej linii, z takim oto określeniem dwóch punktów ciężkości. (1) Pod hasłami logiki i gramatyki poświęcać nadal wiele uwagi zagadnieniom złożoności obliczeniowej, które są dziś fundamentalne dla informatyki, a także dla metodologii nauk empirycznych, w tym nauk społecznych. (2) Pod hasłem retoryki obejmować zagadnienia komunikacji społecznej, zarówno w wymiarze merytorycznym jak i metodologicznym.

Co do retoryki, proponuję zastanowić się chwilę nad jej współczesną szansą, czerpiąc z lekcji historycznej, jakiej dostarcza doniosła jej rola u świtu nowożytności, w wieku XV. Tak pisze o tym Stefan Swieżawski:

> »Problem retoryki w zachodzących wówczas przemianach zajmuje w jakimś sensie pozycję centralną. Liczne i wielorakie więzy wiążą tę dyscyplinę z licznymi dziedzinami myśli i życia praktycznego. [...] Ma to oczywiście wyraźny oddźwięk i w filozofii; wśród należących do niej dyscyplin coraz bardziej dominujące znaczenie zyskuje etyka, wraz z retoryką jako jednym z najsprawniejszych jej narzędzi. Oczywiste są bliskie i liczne związki etyki z prawem, a zarazem i całej rozbudowanej teorii i praktyki prawniczej właśnie z retoryką; ona to nadaje swoistego piętna ówczesnym studiom prawniczym. [...]
>
> Retoryka była wówczas jakby na równi teoretyczna i praktyczna i dlatego słuszna jest teza, że w tym okresie – a w każdym razie we wczesnym humanizmie – polityka utożsamiała się z retoryką, bedącą podstawowym elementem w tych kontaktach między ludźmi, w których dokonują się istotne decyzje polityczne. W tym wczesnym, „obywatelskim" humanizmie włoskim cała odnowiona filologia, z retoryką na czele ma służyć polityce, która ze swej strony zostaje (przynajmniej w zamierzeniu i w teorii) podporządkowana etyce.« (*Dzieje filozofii europejskiej XV wieku.* [tom] *2.* Wiedza. ATK – „Collectanea Theologica", Warszawa 1974, s. 236 n.)
>
> Warto jeszcze przypomnieć, że w owym czasie pojawiły się propozycje tak szerokiego pojmowania logiki, że retoryka stawała się jej częścią. Mamy to udokumentowane w tytule jednego z wykładów Gersona: *Lectio de duplici logica, quarum una est pro speculativis scientiis, alia pro moralibus, quae rhetorica nominatur.* (Wykład o dwojakiej logice, z których jedna jest przeznaczona dla nauk teoretycznych, a druga, zwana retoryką, dla praktycznych. Op.cit., s. 176.)
>
> Ów wątek „gersonowski" podejmuję w pewien sposób w książce *Logic from a Rhetorical Point of View* (de Gruyter, Berlin etc. 1994), gdzie porównuję rachunkowe czyli algorytmiczne partie logiki z badaniami nad efektywnością komunikacji. Te drugie powinny wyjść dalece poza środki algorytmiczne, co jednak ich nie zwalnia od metodyczności; napięcie między praktycznością oraz tą precyzją, jakiej wzór daje logika, jest głównym wyzwaniem dla współczesnej retoryki.

Takie perspektywy mając na oku, trzeba pomyśleć nad naprawą funkcjonowania *Studies* w dwóch żywotnych punktach.

Jednym z nich jest udział autorów zagranicznych. Pod tym względem budujące były pierwsze lata, kiedy w naszym roczniku publikował Kuno Lorenz z Saarbrücken, Guido Küng z Fryburga szwajcarskiego, Kris Coolsaet i Albert Hoogevijs z Gandawy, Lynne Broughton z Melbourne i inni autorzy z ośrodków, z którymi utrzymywaliśmy kontakty. Odnotowawszy regres tego rodzaju współpracy, trzeba dążyć do jej wznowienia. Innym niedostatkiem jest słaba dystrybucja pisma; paradoksalnie, łatwiej jest o autorów niż o czytelników. Wyjściem rozsądnym, doskonale sprawdzonym w świecie spółek handlowych, byłaby fuzja z inną instytucją i powiązanym z nią środowiskiem. Instytucją dostatecznie mocną, żeby poprawiło to zarazem szanse współpracy zagranicznej oraz recepcję pisma w skali krajowej.

Mogłoby to być któreś z towarzystw naukowych o profilu bliskim tematyce *Studies*, jak Polskie Towarzystwo Filozoficzne czy Polskie Towarzystwo Logiki i Filozofii Nauki. Najbardziej jednak obiecującym kandydatem zdaje się być Komitet Nauk Filozoficznych PAN, który dysponuje zapleczem administracyjnym PAN, a jako instytucja z wyborów powszechnych (wśród samodzielnych pracowników nauki) reprezentuje stosunkowo szerokie środowisko. Od lat w KNF nurtuje myśl wydawania pisma filozoficznego w języku angielskim, nie może się jednak zrealizować z braku środków finansowych szacowanych w takiej ilości, jaka wydaje się niezbędna autorom projektu. Jest to jednak oszacowanie mocno zawyżone i stąd skutkujące przedwczesnym pesymizmem. Przeciw takiemu pesymizmowi przemawiają doświadczenia naszego pisma.

Tu niezbędna jest garść detali technicznych, żeby projekt, który rozsnuwam nie był tworem papierowym. Przy założeniu, że zespół redakcyjny pracuje społecznie, wspaniałomyślni autorzy nie oczekują honorariów, a honorowych recenzentów zapewni KNF odwołując się do nastawienia społecznego swych członków (co już się w KNF praktykuje), koszty zaś konsultacji językowej i typografii pozostaną na poziomie białostockim, to nakład 200 egzemplarzy można uzyskać za kwotę 4000 zł, którą to kwotę pismo corocznie uzyskuje z KBN na wniosek Katedry Logiki, Informatyki i Filozofii Nauki Uniwersytetu w Białymstoku. Kwotę tę łatwo podwoić, przy zamierzeniach ambitniejszych (np. wydawanie tomu co pół roku) dzięki środkom na badania statutowe tejże Katedry i jednostek z nią współpracujących oraz wprowadzeniu tej pozycji do budżetu KNF PAN.

Fuzję od strony osobowej można by realizować w ten sposób, że aktywna rola w polityce edytorskiej przypadnie Sekcji Wydawniczej KNF, której przewodniczący zostałby w naturalny sposób przewodniczącym Rady Naukowej *Studies*, a Komitet Redakcyjny powstałby z mianowania KNF, przy wynikającym z natury rzeczy repektowaniu kompetencji, doświadczenia i zasług dotychczasowego składu redakcji.

Jedyne, jakie może się nasunąć zastrzeżenie, to okoliczność, że pismo nie powinno stracić wypracowanego dotychczas profilu, a ten nie ogarnia całokształtu dyscyplin i zagadnień filozoficznych reprezentowanych w KNF. Zróbmy zatem rejestr tego, co z owego całokształtu da się w profilu *Studies* zmieścić: logika, metodologia nauk, filozofia nauki, ontologia formalna, teoria poznania, filozofia umysłu, filozofia języka i komunikacji, duża część filozofii społecznej i politycznej (przykładem tom 18 poświęcony F. Hayekowi), a także filozofia prawa i etyka (por. wyżej uwagi o retoryce wczesnego Renesansu). Nie jest to całość filozofii. Czy jest to obszar zbyt mały, żeby KNF mógł firmować *Studies*? Na moje wyczucie, nie jest to mało, ale ostatecznie niech rzecz ocenią arbitrzy bardziej bezstronni.

*
* *

Na koniec, parę słów o historii przedsięwzięcia. Jak wspomniałem, pomysł zrodził się w roku 1979, kiedy na wydawanie nowego tytułu akademickiego trzeba było pozwolenia Ministerstwa, a jeśli by się o nie wystąpiło, to trudno by nawet liczyć na odpowiedź odmowną; wniosek utonąłby w ministerialnej szufladzie. Uczelnia miała jednak prawo wydawania zeszytów naukowych; rzecz była więc w tym, żeby zmieścić się w tej

formule, ale miast ją ukazywać, przesłonić ją tytułem *Studies in Logic, Grammar and Rhetoric*. Ówczesny szef wydawnictw uczelnianych chwycił w lot intencję (rzekłszy: „rozumiem, chodzi tylko o nowy projekt graficzny okładki").

Pismo było pomyślane jako ukazujące się w językach kongresowych. Istotnie, ukazywały się w nim teksty francuskie i niemieckie, ale przytłaczająca jest przewaga angielskiego. O współpracy z autorami zagranicznymi już wspominałem. Był to wynik ówczesnych kontaktów zagranicznych, z jednej strony moich, z drugiej zaś grupy Mizara (projekt z dziedziny automatycznego dowodzenia twierdzeń) zainicjowanej i kierowanej przez Andrzeja Trybulca. Obecnie Mizar ma własny organ międzynarodowy przyciągający licznych autorów zagranicznych, a moje wysiłki w tym kierunku wchłonęło zainicjowane przed kilku laty pismo elektroniczne po angielsku *Mathesis Universalis*. Tak więc, są tu obecnie zaniedbania, które trzeba będzie odrabiać w kolejnej fazie pisma.

Pierwszy numer rocznika wydaliśmy w 1980 wspólnie z Jerzym Kopanią, wtedy adiunktem w Zakładzie Logiki Filli UW w Białymstoku, dziś profesorem. Następne wydawał on sam, a od roku 1990 przejęła redakcję Halina Święczkowska, adiunkt w tymże Zakładzie; pełni ona funkcję redaktora naczelnego do dziś, dysponując od paru lat dodatkowymi możliwościami wydawniczymi jako profesor wysoce aktywny na Wydziale Prawa UwB, gdzie zajmuje się zagadnieniami komunikacji społecznej.

Znacząca okoliczność, powiązanie treści pisma z kolejnymi Warsztatami Logiki, Informatyki i Filozofii Nauki, była już wspomniana. Zapewnia to pismu wkład najlepszych specjalistów z całej Polski, takich bowiem pozyskiwano dla Warsztatów, zamawiając u nich odczyty za temat zaplanowany przez organizatorów. To gwarantuje wysoką jakość, odciążając od części wysiłków recenzyjnych; opiniuje się natomiast prace młodszych autorów oraz poddaje się ocenie jednego recenzenta całość numeru. Od roku 2001 (nr 17) ukazuje się równolegle wersja internetowa pod adresem: *logika.uwb.edu.pl/studies/*, realizowana technicznie przez Mariusza Giero z Katedry Logiki, Informatyki i Filozofii Nauki UwB, kierowanej obecnie przez Kazimierza Trzęsickiego.

Próbując oszacować perspektywy rozwoju, warto zwrócić uwagę na okoliczność, że periodyk filozoficzny powstał, i przetrwał, prawie ćwierć wieku, w miejscu, w którym nie ma dotąd wydziału ani instytutu filozofii, a miasto Białystok, przy wszystkich swoich urokach i zasługach, nie figuruje na mapie kraju jako znaczący ośrodek aktywności filozoficznej. Brak więc było naturalnego zaplecza, jakim dysponują inne periodyki filozoficzne. Jeśli udało się pismu żyć i rozwijać w tej rozrzedzonej jakby atmosferze, tym większe będą szanse, jeśli doszłoby do fuzji z jakimś zaawansowanym ośrodkiem filozoficznym. Dobrze jest mieć zasługę przetrwania w surowych warunkach, ale nie można z tego czynić celu samego w sobie. A więc, *caeterum censeo, „Studia" reformanda esse.*

ANNIVERSARIAL REMEMBERINGS AND PROJECTS ON THE OCCASION OF THE TWENTEENTH VOLUME *STUDIES IN LOGIC, GRAMMAR AND RHETORIC*

Abstract

Exceptionally, the text being here summarized is written in Polish as its author (not surprisingly) is more flexible in his native langauage, while such a flexibilty is required to impart the text argumentative power in a discussion among Polish philosophers. I mean the discussion concerning future career of the yearly *Studies in Logic, Grammar and Rhetoric*, edited by a team of logicians and philosophers at the University of Białystok, Poland.

The title refers to the medieval Trivium as consisting of Logic, Grammar and Rhetoric. The idea of the journal's scope might have been expressed in terms of semiotics, since these three disciplines, once called *scientiae sermocinales*, approximately cover the field of what is called semiotical research (including logic). However, there are some reasons to be sceptical about any new discipline which, like semiotics, pretends to be highly interdisciplinary (as nowadays also cognitive science does). For, such a claim results in establishing new academic institutes, societies, journals, etc. which duplicate the work of corresponding (though less 'interdisciplinary') traditional institutions, the latter however having their research methods decidedly better defined and developed. Thus, as the author of the project, I preferred to stick to the old venerable triad, in which rhetoric is to be broadly conceived as a theory of communication, related to vital social issues.

The periodical is being published since 1980, roughly one issue every year, in foreign (with respect to Polish) languages, mainly in English. This, obviously, advances collaboration with collegaues abroad. Since 1997 the periodical appears in a new editorial form, now also in Web version (logika.uwb.edu.pl/studies/). During that time, its content to some extent consisted of the invited papers having been submitted to the Polish *Workshops of Logic, Informatics and Philosophy of Science*, devoted to the issues of computability and computational complexity. These are organized every year since 1997 by the Department of Logic, Informatics and Philosophy of Science (University of Białystok) on the occasion of relevant anniversaries (the birth centenaries of Emil Post, then of John von Neumann, etc). This nexus of the journal with conferences on computability has considerably influenced the content of *Studies* in the last few years.

When considering prospects of the journal, I suggest to try a merger with some significant philosophical institution, highly representative of philosophy in Poland. Thus the journal would be edited by a new team composed of the old one and representatives of the new collaborating institution. This should extend the circle of contributors and that of readers, increase financial means, facilitate refereeing, and advance propagation.

Witold Marciszewski
University of Białystok

CHALLENGES FOR THE LOGIC OF SOCIAL RESEARCH: TO GRASP RATIONALITY, TO DEAL WITH COMPLEXITY

The two Herculean tasks – to define rationality, as a basis of social order, and to tackle complexity of social phenomena – require harnessing potent forces and resources of logic. One needs 'basic logic', that is, logical calculi to have rules of correct reasonings, then methodological reflexion on the use of mathematical models in social sciences, at last some additions to basic logic. The last involve theoretical computer science to judge the power of algorithms used in modelling, and a study of practical reasoning in social interactions; such a study is provided by the theory of games and decision-making. All that jointly deserves to be called *the logic of social research*[1].

1. Logic, mathematical modelling, and artificial societies

The theory of logic, insofar as we attain to it, is the vision and the attainment of that Reasonableless for the sake of which the Heavens and the Earth have been created. This enthusiastic belief in logic as expressed by Charles Sanders Peirce, is what Evert W. Beth referred to with the motto of his seminal study *Semantic Entailment and Formal Derivability* (Amsterdam, 1955).

In that study Beth succeeded in grasping an essential feature of '*that reasonabless*' which one finds in first-order logic and the accompanying metalogical reflexion. This is just a part of the idea of rationality, but a part significant enough to be taken it as the starting point of discussion.

[1] This paper was supported with financial means of the State [Polish] Committee for Scientific Research as research project no. 2 H01A 030 25: *Nierozstrzygalność i algorytmiczna niedostępność w naukach społecznych* (Undecidability and Algorithmic Intractability in Social Sciences), run in 2003-2006.

ISBN 83-7431-015-4 ISSN 0860-150X

Beth's study, inspired by some Gentzen's idea, offers a very important logical contribution to the notion of rationality. His predicate logic system called *semantic tableaux* is much worth attention for it represents the most algorithmic approach in solving the problem about an inference whether it is logically correct. It is not the only system of this kind but the one which historically was the first in a chain of similar systems, and is a fitting example to represent this whole chain. The problem of whether a formula of predicate logic is logically valid can be solved in different ways. Some of them involve the guessing of premisses from which the formula in question could be deduced; this gives us opportunity to show invention, but does not guarantee success. In other strategies, to which Beth's method belongs, the process of checking validity is guided by a set of algorithmic rules to obtain a result, provided that it is obtainable at all (this not always is the case because of the undecidability of logic). This algorithmic strategy is more rational than the other one, in the sense of being more efficient.

A new level of rationality is attained with metalogical reflexion concerning theoretical tools which one uses in mathematical modelling. A researcher who is aware of metamathematical properties of a theory used for modelling should be acknowledged as more rational than the one who lacks such awareness. The well-known story about Frege who recognized inconsistency of his system of logic after Russell's critique (the antinomy of the set of sets not being their own elements) exemplifies such degrees of rationality; Frege's attitude towards his system was more rational after his learning about the flaw[2]. Completeness of a theory is another subject of such reflexion, still other ones are decidability and computational tractability (decidability in practice).

Such metalogical assessments add new points to the postulate of logical correctness of reasonings. Suppose, there are two researchers in an empirical science. One of them adopts a mathematical model provided by a naive non-formalized theory, whose decidability has not been thus far investigated, while the other employs the same (as to the content) theory stated in a formalized form and enjoying a proof either of decidability or of undecidability. The latter should be appreciated as acting in a more rational way as being more conscious of his tools and methods. Another comparison can be drawn between two theories, one of them being computationally tractable, while the other not, the comparison resulting to the advantage of the former.

The use of mathematical theories to provide models in an algorithmic form, necessary for computer simulation, faces us with two questions, one belonging more to mathematical logic, the other more to computer science. The questions are as follows: (1) the issue of algorithmic decidablity, (2) the issue of computational tractability. The first is concerned with what can be computed in principle, while the second – with what can be computed in practice (i.e., with respect to available resources of time, memory, etc).

These logical explanations of what rationality consists in, are fundamental for building Artificial Intelligence not less than for understanding the natural intelligence of humans and animals. Any inferences we expect, e.g., from agent programs, should conform to the standards of valid deduction and to metatheoretical requirements, before we address the issue, possibly, with more specialized theories of reasoning.

The domain of AI has its natural extension in what is called Artificial Societies. AS researchers carry computational simulations of social interactions occurring among artificially intelligent agents. How such simulations are related to logic, can be learnt when reading the quarterly *Journal of Artificial Societies and Social Simulation*, abbreviated as JASSS. An instructive example is found in a recent paper by Maria Fasli, entitled with the question *Formal Systems and Agent-Based Social Simulation Equals Null?*. The author answers in the negative, while asserting that there is a considerable common part of the areas in question, and much space for collaboration. A wide class of logical systems is taken into account as those (to quote the paper) "ranging from classical propositional logic and predicate logic to modal logic, dynamic logic and higher order logics"[3].

This example demonstrates that artificial societies is logicians' business as well. Another argument concerning the use of logic takes into account the usual mathematical modelling in social, including economical, research, its procedures being viewed from a logical point of view. As examples of social phenomena mathematically modelled one can mention population dynamics, group interactions, political transitions, evolutionary economics, urbanisation, etc.[4] As much as there appear conflicting interests among agents in such processes, a formal model to render the situation can be ta-

[2] Though Frege did not intentionally work in a paradigm of modelling, any logical system, including that of him, can perfectly serve as a model of mental processes of reasoning.

[3] See Volume 7, Issue 4, 31-Oct-2004; jasss.soc.surrey.ac.uk/JASSS.html.

[4] Cp. Weidlich, Wolfgang: *Sociodynamics: a Systematic Approach to Mathematical Modelling in the Social Sciences*, Harwood Academic Publishers: Amsterdam, 2000; reviewed by Glen Lesins in JASSS, the issue mentioned above.

ken from von Neumann's and Morgenstern's mathematical theory of games which provides us with the fitting model called prisoner's dilemma.

Algorithmic decidability and computational tractability are thoroughly considered by the physicist Stephen Wolfram (especially in his widely discussed book *A New Kind of Science*, 2002) with respect to empirical theories of physics and biology. Wolfram uses the computational model of cellular automata to examine decidability and tractability of some theories in natural science.

The mentioned book by Wolfram, though being a considerable academic event in the US and other Western countries, hardly received (as far as I know) any attention either from logicians or from social scientists in Poland. This seems to be a more general symptom of one's indifference to the issues of decidability of empirical theories. In spite of the enormous achievements of Polish logicians, especially in the period between the two world wars, nowadays their presence at the area of such vital applications of logic is hardly felt.

The realizing of this fact motivated the present author to arrange meetings of logicians and social scientists from main centres of these disciplines in Poland. Such meetings, first, should have yielded a survey of interests, projects and skills, and then should have given an idea of how to enrich future research in Poland with the (neglected thus far) issues of computability. Let me report on two such workshops, as their content closely coincides with the subject of this paper. The two workshops in question belonged to a series of annual meetings devoted to the issues of decidability and computability in a historical perspective[5].

2. A case of social computability: spontaneous order vs central planning

In 2002 there took place the Workshop: *In F. Hayek's 10th Death Anniversary – Free Market as Information-Processing System. On the Problem*

[5] The series of conferences under the name *Workshops in Logic, Informatics and Philosophy of Science* was organized by the Committee for Philosophy of Polish Academy of Sciences in collaboration with the Department of Logic, Informatics and Philosophy of Science, University of Białystok, abbreviated below as DLIPS-UwB, and other partners, including Centre of Market Psychology in Leon Koźmiński Academy of Entrepreneurship and Management, Warsaw, abbreviated below as CMP-LKA, and Higher School of Public Administration in Białystok. A considerable set of Workshop papers, translated into English, is published in this and in the previous issue of this journal. The Workshops have Web documentation at the site www.calculemus.org/.

of Algorithmization in Social Research. Here are the titles (ad hoc translated into English) of invited contributions (for abbreviations of school names – see footnote 5).

1. Hayek on free competition (Justyna Miklaszewska, Jagiellonian University, abr. UJ)
2. Hayek: The idea of self-organization and the critique of constructivist utopia (Wiesław Banach, Adam Mickiewicz University, Poznań)
3. Spontaneous order in social philosophy: from B. Mandeville to F. A. Hayek (Miłowit Kuniński, UJ)
4. The role of instability in dynamic systems (Michał Tempczyk, Polish Academy of Sciences)
5. The stabilizing role of non-linearity (Michał Tempczyk)
6. Chaos in economics (Arkadiusz Orłowski, Informatics Department in Warsaw Agricultural University)
7. Econophysics – a new paradigm? (Arkadiusz Orłowski)
8. On mathematical models used at stock exchange (Bolesław Borkowski and Arkadiusz Orłowski)
9. Market efficiency and the behavioral finance theory (Piotr Zielonka CMP-LKA)
10. Computational tractability and the physics of information (Arkadiusz Orłowski)
11. From genealogy of mathematical economics: Walras, Pareto. Lange (Anna Zalewska, DLIPS-UwB)
12. L. Savage's mathematical theory of decision-making (Dariusz Surowik, DLIPS-UwB)
13. What the central planning cannot do (Andrzej Malec, DLIPS-UwB)
14. The problem of computational tractability of social structures – in the example of the computational power of free market (Witold Marciszewski, DLIPS-UwB).

Each of these items in a way contributes to the explaining of the key notions of this essay as appearing in the title: 'to grasp rationality, to deal with complexity'. Let me hint at connexions.

Papers 1–3 introduce some crucial ideas due to Hayek, while the remaining ones provide a conceptual apparatus to develop these ideas. *Free competition, self-organization, spontaneous order*, discussed in those papers, are phenomena which Hayek, especially in his polemics with Oskar Lange (cp. items 11 and 14), considered with respect to the issues of complexity and computability. Contrary to Lange, Hayek claimed the advantage of free market over central planning (cp. item 13) with respect to power of computation.

Items 4–7 are to direct attention to the theory of complex dynamic systems as explaining some sources of unpredictability of social processes. This is an approach to complexity which is complementary to that addressing the notion of algorithmic complexity.

Items 8–13 deal with models of economic processes, treated as that representation of social processes which best exemplifies problems of applicability of mathematical models and algorithms. Item 11 sheds light on a methodological aspect of socialist economics. Oskar Lange, who was a strenuous advocate for mathematical models and computerized calculations in socialist economics, referred to the school of economics initiated by Walras and Pareto; it is the context in which his ideas of socialist economy become more conceivable. When taking into account extreme simplifications in Walras and Pareto models for economy, one fully appreciates the role of the concept of computational tractability. Even if socialist economy had a consistent model, the real economic life, when compared with such models displays complexity which exceeds any algoritmic capabilities and computational resources necessary for socialistic central planning. This was intuitively grasped by Ludwig von Mises and Friedrich Hayek, while a precise formulation of their ideas is nowadays possible due to the theory of computational complexity; this was the point of the closing item 14.

To put the thing in a nutshell: free market is a computational device which for at least three reasons is more efficient than central planning; this is due to the use of computing which is distributive, interactive and analog.

Owing to *distributivity*, which means that each market agent processes the data necessary for his own business alone, the volume of data to be processed is enormously smaller (than in central planning). Thus computational complexity does not exceed computational capacities.

Interactivity means that a system is able to learn data-processing from its feed-back interactions with the environment (as entrepreneurs do). It does not need to be fully preprogrammed, what would require unimaginably more memory space and computing time. For, instead of programs and databases preparing a system to meet all possible situations, only a program for self-learning is needed.

Analog computation proves more efficient since it does not require rendering all the process with symbols. This spares time and makes data processing possible also in these situations in which impulses cannot be strictly measured, being only approximately felt by individual receivers. Often such approximations are sufficient for rational decision-making, being at the same time dramatically swifter and more economic than any symbolized and digitalized data-processing. And this is that much sensitive way free market agents run their bussiness, in contradistinction to central planning officials who are unable to handle any data if not recorded in symbols.

A social order arranged according to the above three principles of data--processing usually emerges spontaneously as a result of long evolution, that is, a trial and error self-organizing process. This is what Hayek called *spontaneous order*, and saw it in free market, the development of language, democracy, self-governance, etc. Usually, spontaneous order involves an ethical code that outlines acceptable behavior within the unit or group.

As testified by the fitting forecasts of liberal thinkers (Friedrich Hayek, Karl Popper, Ralf Dahrendorf, Richard Pipes et al.), the concept of spontaneous order has a considerable predictive force. In particular, it implies the collapse of those systems which are extremely hostile to spontaneous order, and introduce instead artificially constructed centralized systems which have no chances when faced with complexity of real life. This was the basis of prediction which Richard Pipes made regarding the Soviet Union. In his autobiography *Memoirs of a Non-Belonger*, 2003, (Chapter 2, Section on Historical Revisionism) Pipes claims that only liberalism, as decentralizing decision-making, is able to handle the complexity of the contemporary world. Pipes, an outstanding historian of Russia and the Soviet Union, is no theorist of computational complexity. Nevertheless, on the basis of historical experience, he arrives at the view like Hayek's that the tractability of complexity is the main political challenge. This idea belongs to the very core of liberalism, on the same footing as the respect for freedom and human rights.

3. The theory of games and computational tractability

In 2003 the Workshop devoted to the above theme was entitled *Computational Power for Social Research. Von Neumann's Ideas – in the Centenary of His Birth.* On account of the Centenary celebration, the program included also talks on some von Neumann's merits beyond the scope of social research, as his contributions to quantum theory and other branches of physics (A. Orłowski), meteorology in the context of Chaos (M. Tempczyk), metamathematics (R. Murawski, J. Pogonowski). The main stream of Workshop involved the theory of games with utility theory, cellular automata as models of social phenomena, and mathematical modelling related to these both subjects. Here are the titles.

1. Evolution of the concept of utility (Tadeusz Tyszka, CMP-LKA)
2. A brief history of Game Theory (Jaideep Roy, CMP-LKA and Indian Statistical Institute, New Delhi)

3. The concept of utility. Should it be revisited? (Janusz Grzelak, University of Warsaw, abbr. UW)
4. The theory of social choice – from Arrow to the current state (Grzegorz Lissowski, UW)
5. The Nice voting system vs. Convention System. An application of power index (Mikołaj Jasiński, UW)
6. An empirical example of limited rationality (Honorata Sosnowska, Warsaw School of Economics)
7. The modelling of the influence of two basic inclinations of investors upon stock indexes (Piotr Zielonka, CMP-LKA)
8. Temporal logic and game theory (Dariusz Surowik, DLIPS-UwB)
9. Dynamic minimalism: the role of computer simulations with cellular automata in studying social processes (Andrzej Nowak, UW and Florida Atlantic University)
10. Social influence modelled in cellular automata: basic research and applications to economic processes in Poland (Andrzej Nowak)
11. Theory of cellular automata in simulating social processes (Katarzyna Zbieć, DLIPS-UwB)
12. Some applications of cellular automata in constructing self-learning systems (Paweł Borkowski, WSP, Częstochowa)
13. The concept of rational action with Max Weber (Radosław Oryszczyszyn, UwB)
14. Empirical aspects of computability theory. That is: does the Universe compute better than (thoughtless) man? (Jerzy Mycka, Maria Skłodowska-Curie University, Lublin)
15. On how the methodology of social sciences meets the issues of computational complexity (Witold Marciszewski, DLIPS-UwB).

The main stream of the conference was concerned with (A) game and utility theories (items 1–8), the second to it was (B) the issue of cellular automata, for short CA (items 9–12), the third – (C) the issue of rationality as escaping algorithmicity (items 13–15).

Streams A and B are closely interconnected not only because of the fact that both game theory and CA theory derive from John von Neumann's works (an advantageous fact in celebrating his Centenary), but also for their being recently combined in social modelling. This combining is nicely exemplified in Andrzej Nowak's own research as reported by him in 9 and 10; some results of other authors were mentioned in communication 11.

CA theory is surprisingly fit to be combined with such paradigmatical cases of game theory as the Prisoner's Dilemma (for that Dilemma see item 3, discussed by J. Grzelak in this issue of *Studies*). The cells in the space of a game in natural way can be regarded as players, while the dilemma in question consists in conflicting interests of theirs. Rules of behaviour in such

a dilemma can be easily stated as strategies being chosen by individual cells which react to impulses from the envirnoment. In those social processs which are modelled as iterated games (i.e., played many times), the players can learn from observing results of strategies adopted by them and those adopted by partners[6]. What is specially interesting in studying social evolution, it is the phenomenon of cooperative strategies. These emerge either from social-oriented (altruistic) reactions, or even from one's self-interest when collaboration proves more advantageous for an egoist. CA rules display that astonishing property that even when being very simple, after some number of steps they lead to unexpected results, thus proving their predictive value. In such a context there arise some issues of decidability, for instance, whether the following problem is decidable: suppose that from a given point of time a strategy established itself as dominating: will it remain dominating forever? (Cp. Grim 1997 in Literature).

The third stream of the Workshop, represented by items 13–15, was rather marked than fully developed. Let it be advanced a bit further now, in the next Section. Item 13 from the list of topics yields a convenient starting point.

4. Value-oriented rationality as a basis for social predictions

The famous among sociologists Max Weber's distinction between what is goal-oriented rational (zweckrational) and value-oriented rational (wertrational) requires elucidation. Weber did not succeed in explaining the latter. Moreover, it got blurred with Weber's sermon claiming that social researchers should refrain from any valuations, thus forbidding them to resort to one of those acknowledged by him sorts of rationality. The thing grows even dimmer when Weber himself uses some phrases in an evaluating tone, e.g. when contrasting as opposities the eternal *auri sacra fames* (abominable greed for gold) and the rational motive of gain, the latter seeing as characteristic of modern capitalism and apparently more valuable than that old greed. Thus it is up to the present writer to offer some explanations, while appreciating Weber's merits for calling the issue to mind.

This calling to mind is topical and timely after some successful political predictions, ones of the great consequence, have been made on the basis of axiological assumptions. For instance, Richar Pipes – the renowned historian

[6] Compare the Workshop on *Agents that Learn from Other Agents* held as part of the 1995 *International Machine Learning Conference*. See www.cs.wisc.edu/shavlik/ml95w1/.

of Russia and Ronald Reagan's influential adviser – foretold the collapse of the Soviet Union after having realized the extent of moral ruin of the Soviet system and Soviet people, such a ruin resulting in economic, political and material collapse.

This Pipes' idea is by no means new. It goes back to the Old Testament prophets, as well as the old Romans who believed that moral virtues supported political power; its visible traces can be found in Polish patriotic literature (Jan Kochanowski, Adam Mickiewicz), and in nowadays folk sociology. However, that an idea is old and common among people does not diminish its chances to be true.

On the other hand, the fact that someone inferred from it a prognosis that proved true does not warrant the truth of the premiss. I do not mean to argue that the idea of value-oriented rationality is right; this would exceed the scope of the present paper. Instead, I offer something like a test so that anybody interested might to get aware of his own opinion. One of the views resulting from the test should help in clarifying the notion of value-oriented rationality; those who do not accept that view may device arguments against it. The test runs as follows. Suppose you have two options.

A. To greatly help a honest person being in need at a little expense on your part.

B. To ruin somebody at the same expense as above, with no gain for you except a pleasure of dominating through inflicting pain.

Suppose the only motive of choice you consider is doing what is more right. Now decide which option is regarded more right by you, and express the choice in a verbal Statement, say (as I may guess):

[S] A is more right than B.

Now there is time for the next decision. Do you regard S as a genuine sentence in declarative mood, that is, *a mood that represents the act or state as an objective fact* (as defined in grammar)?

If you answer in the afirmative (what, possibly, is not the case if you are an orthodox neopositivist), this means that you assert S as true. Think once more whether you like regarding S as true, because if you stick to that, you have to acknowledge the state of something being more right than something else 'as an objective fact' (according to the quoted definition); and this, by no means, is either psychological or physical fact, it is rather a 'metaphysical fact'. So be cautious, if you are afraid of engaging yourself into metaphysics. But, think again. If you withdraw your opinion of S as being true, what do you offer in return? You should have a positive suggestion, including the

explanation what a grammatical kind of sentence would be disguised in the indicative form of S.

Now, the sequel is addressed to those who accept S as true, thus qualifying it as an indicative sentence. If so, then there may be a logical following of an empirically testable sentence (as being indicative too) from such propositions as S; let us call them *estimates*, be them moral, esthetic, or alike. Thus an empirical prediction may logically follow from a set of premises including estimates. In order to create such a nexus between empirical propositions and estimates, we need some principles of the kind like the following Principle of Axiology:

[PAx] *In a long run, any social system to survive needs axiological foundations, that is a set of non-empty basic axiological concepts to be used in estimates.*

They are basic in the sense that they do not derive from any other concepts but are primitive, like concepts occurring in axioms of a deductive system. Such are the notions of justice, freedom, credibility, loyalty, dignity. That they are so basic as if they were inborn to human minds, could be observed in the Ukrainian protest against iniquities of the political establishment such as falsified results of election in 2004. Similar obviousnesses were felt by those Polish, Russian or Czech people who earlier resisted the communist regime in the period of its absolute dominance.

Such events yield a direct exemplification of PAx. The total lack of axiological foundations makes totalitarian regimes unstable – contrary to the cynical opinion (of dictators and their followers) that brute power which does not respect any moral principles is the most reliable means of lasting dominance and stability.

PAx is just that principle which for people like Richard Pipes became one of two main premisses in forecasting the collapse of the Soviet Union block. The other premiss, already mentioned above, was to the effect that the extreme centralization of economy and politics is a ground of weakness bringing about the collapse of a system.

Thus we obtain an evidence that there are principles, those like PAx, which may function as premisses in reasonings about social affaires. At the same time, we obtain an example what value-oriented rationality may be like (thus filling a gap in Weber's exposition). And it is also the case of a challenge facing the logic of social research, as signalled in this paper's title.

That logic should explain the following question: is it possible for a social theory built on principles like PAx to have a mathematical model in algorithmic form? Should the answer be in the negative, we are to choose between two alternatives. Either there is a mathematical model but not algorithmic one, that is, one that does not imply algorithmic computability (algorithmic decidability), or there is no mathematical model. The former option is one that would engage us into a new, unconventional and little known field called hypercomputation, that is computation beyond the limits of Turing machine; this is the subject dealt with in a contribution to von Neumann Workshop (see item 14 as listed in Section 3, and the corresponding contribution to this issue). The latter option (no mathematics) might belong to what Weber called *understanding sociology*, that is, a discipline which explains social phenomena by resorting to the notion of rationality; in this case it would be value-oriented rationality (this, however, is a matter for further discussion, as other interpretations of the understanding sociology programme, reconciling it with a mathematical approach, may be considered).

The challenge is here just recorded, not responded. However, in the moment such a mere recording is a step forward – the more needed, the less its need is obvious for community of scholars.

5. Does brain's complexity surpass that of human-devised algorithms?

Janusz Grzelak in the contribution to this volume ends his discussion on the concept of utility with an optimistic statement which runs as follows.

> The concept of utility is growing more complex, but the measurement possibilities of contemporary psychology appear to match its level of complexity.

However, the unavoidability of principles like PAx (introduced in previous Section) appears to hint at the contrary. The complexity of moral drives as permeating societies in some periods of their histories (as in the struggles for moral values against communist regimes) seems to exceed explanatory power of nowadays sociology or social psychology. In a remarkable way, the simple mind of Ronald Reagan (derided for that simplicity by intellectuals) proved more understanding and foreseeing than scientifically looking theories of numerous Sovietologists. Here is a fitting characterization of the latter given by Richard Pipes[7].

[7] See an excerpt from Richard Pipes's *Memoirs of a Non-Belonger* (Yale University Press 2003) reprinted at the website http://hnn.us/articles/1836.html.

> Insisting that moral judgments have no place in science (and they considered themselves scientists) the Sovietologists treated societies as if they were mechanisms. One of their basic premises held that all societies performed the same "functions", even if in different ways, on which grounds they interpreted in familiar terms all those features of the communist regime which to a mind untutored in social science appeared outlandish. One such "expert", for example, found no significant difference between the way New Haven was administered and any city of similar size in the Soviet Union. (2) The net result of this methodology was to depict communist societies as not fundamentally different from democratic ones: a conclusion that reinforced the policy recommendation that we could and should come to terms with them.

Moral judgments in science, alien to the content of Sovietological theories, seem inopportune also to some other scholars for reasons which are purely methodological. May be, they are right when refuse to accept as scientific something which is not capable of being measured and mathematically modelled. But then they should acknowledge the superiority of a non-scientific mind, like Reagan's, whose intuition in foreseeing social processes has turned out more powerful than predictive force of scientific theories. Thus it seems that if the theorists do not like appearing useless, they should reconsider their refusal of including value-oriented rationality into the paradigm of social sciences. Such a turn might be justified in the following way.

There may be a cognitive power in some brains which surpasses the scope of what can be recorded in a language and then expressed in a theory. Therefore theories have to resort to simplifications. This can be a good strategy, provided that simplifications get overcome if necessary and justice is done to greater complexity. Such a process in which complexity of a theory increases in order to match the complexity of the phenomena under study, can be traced in the story (in this volume) *Historic and contemporary controversies on the concept of utility* by Joanna Domurat and Tadeusz Tyszka. First, the notion of rationality in John von Neumann's and Oskar Morgenstern's game theory is in that paper described, then we learn that one of its simplifications was remedied by Savage's definition of subjective probality, then another one by D. Kanehman and A. Tversky's prospect theory; the last amends the previous game-theoretical simplifications with taking into account reference points (e.g., some previous experiences) which influence one's utilities.

This story is thought-provoking since it reveals how preverbal insights must precede improvements of a theory. Before a new or improved theory is stated, the mind has to realize a discrepancy between the old theory and a newly observed feature of reality, while the old theory does not render that feature in its verbal statements, e.g., its axioms. Thus the question arises about superiority of mind/brain pretheoretical processes over those going on with the use of a theory.

To address the problem, let me start from quoting an inspiring hypothesis going back to von Neumann[8].

> It is only proper to realize that language is largely a historical accident. The basic human languages are traditionally transmitted to us in various forms, but their very multiplicity proves that there is nothing absolute and necessary about them. Just as languages like Greek or Sanscrit are historical facts and not absolute logical necessities, it is only reasonable to assume that logic and mathematics are similarly historical, accidental forms of expression. They may have essential variants, i.e., they may exist in other forms than the ones to which we are accustomed. Indeed, the nature of the central nervous system and of the message systems that it transmits indicate positively that this is so. [...] Thus logic and mathematics in the central nervous system, when viewed as languages, must structurally be essentially different from those languages to which our common experience refers.

The problem so raised involves four more particular questions.

(1) Is the logic and mathematics of brain subject to restrictions of limitative theorems concerning undecidability (computability in principle)?

(2) If so, is the scope of undecidable problems identical with that constraining the Turing machine?

(3) If not identical, then how should it be defined?

(4) Is there any superiority of the brain over Turing machine embedded in digital electronic computers in respect of computational tractability (practical computability)?

The last problem obtained an answer quite recently. In a contribution entitled *The Code of Mathematics: John von Neumann's "The Computer and the Brain" (1958)* (in the forthcoming book, 2005, by Dirk Baecker (Ed.), *Key Works of Systems Theory*, Opladen: Westdeutscher Verlag) Loet Leydesdorff (Amsterdam School of Communications Research, University of Amsterdam), when summing up some earlier results, writes as follows[9].

> Von Neumann noted that the mechanism of using shorter code may also work the other way round: the material substance may contain mechanisms which are different from our common mathematics, but which are able to solve problems by using shorter code. This prediction has come true. For example, the so-called "traveling salesman problem" which is virtually impossible to solve using normal computation (because it is NP complete), can be solved by using DNA strings in a laboratory setting thanks to the properties of this material (Adleman, 1994; Liu et al., 2000; cf. Ball, 2000). The biochemistry of the system must be understood in addition to the mathematical problem. The recombination of formal and material insights provides us also with new mechanisms for the computation of complex problems. Thus, the mathematics can function as a formal bridge between the special theories that remain otherwise specific.

The notion of short codes is due to Alan Turing. Such a code enables a second machine to imitate the behavior of another, fully coded, machinery. Short codes were developed to make it possible to code more briefly for a machine than its own system would allow. This consists in treating it as if it were a different machine with a more convenient, fuller system of instructions which would allow simpler, less detailed and more straightforward coding.

The message (in the Leydesdorff's passage quoted above) which is crucial for our discussion of game theory as this theory pretends to provide algorithmic models of decision-making and of social processes (competition, collaboration, etc.) runs as follows:

— *There is a NP-complete problem being solvable by a short code in the brain, namely the Travelling Salesman Problem (TSP).*

This premiss should be combined with another one, belonging to the so-called theory of coherence (to be mentioned later).

The above result (italicized) implies that in dealing with complexity of TSP the brain by far surpasses computational capabilities of the digital electronic computer. Here is in order to recall that NP-problems are those

[8] The quoted passage is the last paragraph in the last von Neumann's work *The Computer and the Brain*, Yale University Press, 1958. The book is a record of a lecture series that von Neumann delivered at Yale University in 1956.

[9] The excerpt is copied from the page users.fmg.uva.nl/lleydesdorff/vonneumann/. The items referred to in abbreviations are as follows.
- Adleman, L. M. (1994). "Molecular Computation of Solution to Combinatorial Problems", *Science* 266 (11), 1021-1024.
— Ball, P. (2000). *DNA computer helps travelling salesman*, at http://www.nature.com/nsu/000113/000113-10.html.
— Liu, Q., Wang, L., Frutos, A. G., Condon, A. E., Corn, R. M. and Smith, L. M. (2000). "DNA computing on surfaces", *Nature* 403, 175.
This message was found by me at the last stage of writing the present paper. In an earlier version there was a passage in which I wrote about this von Neumann's hypothesis as one waiting for empirical confirmation. Fortunately, a confirmation is now the case. As for the Travelling Salesman Problem, a lot of information can be found at Web, when using this phrase as an entry.

which are (for electronic computers) uncomputable in practice, and they are so since algorithms needed for their solving work in exponential time, this trait being the measure of problem's complexity. When a problem is called NP-complete this means its belonging to the hardest problems in the NP category, that is (let me repeat), the category of problems whose solving requires exponential algorithms[10].

The discovery of so enormous advantages of brains over computers sheds light on game-theoritical algorithms when compared with capabilities of human intuition, the latter, presumably, being due to the brain's equippment. The bridge between the above TSP issue and game-theoretical models is to be conceived as follows. The process of decision-making involves considering all the possible outcomes of actions (which could be taken by the decision-maker) in order to check which incomes are mutually coherent. This is the MCO problem:

(MCO) *How to Maximize Coherence in the set of all possible Outcomes of the actions considered before choice.*

Thus in models of decision-making we need an algorithm for searching outcomes. For n outcomes the number of possible combinations equals 2^n, say 2^{100} with 100 outcomes (in the worst case), what is no unrealistic situation. A problem requiring such an exponential algorithm for exhaustive search belongs to the NP-complete category. It is the same as the category of TSP. Hence any device capable of solving TSP is capable of solving MCO. Since the brain is a device which solves TSP (according to the result reported above), it can solve MCO as well, while for the electronic computer both problems are in practice unsolvable.

From psychological side, that biological capability should be, plausibly, identified with what people used to call intuition and what proves to be a potent tool to deal with complexity. In turn, that intuition, when concerned with rationality, is what the understading sociology relies on. Thus, *via* game theory and algorithmic complexity theory we return to the old venerable programme, mainly due to Max Weber. When taking his idea in the above way, based on these modern, reliable and flourishing disciplines (complexity theory and neurobiology), we may have chance to build the understanding sociology on solid logico-empirical foundations.

[10] For this terminology see, e.g. db.uw/aterloo.ca/alopez-o/comp-faq/faq.html, or www.nist.gov/dads/HTML/npcomplete.html.

So far only the last of the four problems listed at the start of this Section (after quoting von Neumann's text) was discussed. Even when being treated in a very sketchy way, this item required a considerable space. Much more would require every of the remainig issues, which besides have the drawback that no empirical results, comparable with those concerning the capabilities of brain DNA have been attained so far. However, the very stating of problems, even without hints toward solution, seems to be of use as a small preliminary step. Hopefully, not before long greater and more conclusive steps in that direction will be made.

Literature

The list of Literature below contains items recommended for reading as useful guides to the issues which are discussed in this paper. As for items used for references, they are quoted inside the paper, either in the main text or in footnotes.

Selmer Bringsjord, Michael Zenzen: Toward a Formal Philosophy of Hypercomputation. *Minds and Machines*, Volume 12 Issue 2, May 2002.

Jerzy Brzeziński et al. (Eds.): *Creativity and Consciousness. Philosophical and Psychological Dimensions*, Rodopi, Amsterdam – Atlanta 1993.

Jack Copeland: The Broad Conception of Computation, *American Behavioral Scientist*, 40, 1997, 690-716.

Jack Copeland [2000, Copyright]: *Turing's O-Machines*, www.alanturing.net /turing_archive/pages/Reference%20Articles/referencearticlesindex. html

Jack Copeland: Accelerating Turing Machines. *Minds and Machines*, Volume 12 Issue 2, May 2002.

Jack Copeland: Hypercomputation. *Minds and Machines*, Volume 12, Issue 4, November 2002.

M. L. Hogarth: Non-Turing computers and non-Turing computability. vol. 1, PSA, 1994.

Patrick Grim: The Undecidability of the Spatialized Prisoner's Dilemma, *Theory and Decision* 42, 1997, 53–80.

J. Hartmanis and R. Stearns: On the computational complexity of algorithms, *Transactions of the AMS* 117, 1965, 285–306.

J. Hartmanis: Gödel, von Neumann and the P=?NP problem, *Bulletin of the European Association for Theoretical Computer Science* 38, 1989, 101–107,

David R. Henderson: *Biography of Friedrich August Hayek (1899–1992)*, www.econlib.org/library/Enc/bios/Hayek.html

Andrzej Nowak and Maciej Lewenstein: Modelling social changes with cellular automata. In: Rainer Hegselmann (ed.), *Modelling and Simulation in the Social Science from the Philosophy of Science Point of View*, pages 249–285. Kluwer Academic Publishers, Dordrecht, 1996.

Oracle machine – entry in *Brainy Encyclopedia* www.brainyencyclopedia.com/encyclopedia/o/or/oracle_machine.html

Roger Penrose: *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford Univ. Press, Oxford etc. 1989.

Roger Penrose: *Shadows of the Mind*, Oxford University Press, New York 1994.

Oron Shagrir: Super-tasks, accelerating Turing machines and uncomputability. *Theoretical Computer Science*, Volume 317, Issue 1–3, June 2004.

Hava T. Siegelmann: Neural and Super-Turing Computing. *Minds and Machines*, Volume 13, Issue 1, February 2003.

Alfred Tarski, in collaboration with Andrzej Mostowski and Raphael M. Robinson: *Undecidable Theories*, North Holland, Amsterdam 1968 (Studies in Logic and the Foundation of Mathematics).

Alan Turing: On computable numbers, with an application to the Entscheidungsproblem, *Proc. of the London Math. Society,* Series 2, 42, 1936, 230–265.

Alan Turing: Systems of logic based on ordinals, *Proc. of the London Math. Society*, Series 2, 45, 1939, 161–228. Cytowane jako Turing [1938].

John von Neumann and Oskar Morgenstern: *The Theory of Games and Economic Behavior*, 3rd ed., Princeton University Press, 1953.

John von Neumann: *The Computer and the Brain*, Yale Univ. Press, New Haven 1958.

Max Weber: *Wirtschaft und Gesellschaft. Grundriss der verstehenden Soziologie.* Fünfte, revidierte Auflage. J. C. B. Mohr (Paul Siebeck), Tübingen 1972.

Peter Wegner: Why interaction is more powerful than algorithms. *Communications of the ACM*, v. 40 n. 5, p. 80–91, May 1997.

Stephen Wolfram: Undecidability and Intractability in Theoretical Physics, *Physical Review Letters* 54, 1985, 735–738.

Stephen Wolfram: *A New Kind of Science*, Wolfram Media, Inc., 2002.

James A. Yunker: Post-Lange Market Socialism. An Evaluation of Profit-Oriented Proposals, *Journal of Economic Issues*, September 1995.

[Author's name lacking] *The Socialist Calculation Debate* cepa.newschool.edu/het/essays/paretian/social.htm

Roman Murawski
Adam Mickiewicz University

JOHN VON NEUMANN AND HILBERT'S SCHOOL OF FOUNDATIONS OF MATHEMATICS*

Abstract. The aim of the paper is to describe the main achievements of John von Neumann in the foundations of mathematics and to indicate his connections with Hilbert's School. In particular we shall discuss von Neumann's contributions to the axiomatic set theory, his proof of the consistency of a fragment of the arithmetic of natural numbers and his discovery (independent of Gödel) of the second incompleteness theorem.

1. Introduction

Contacts of John (then still Janos, later Johann) von Neumann with David Hilbert and his school began in the twenties of the 20th century. Being formally a student of the University of Budapest (in fact, he appeared there only to pass exams) he was spending his time in Germany and in Switzerland studying there physics and chemistry as well as visiting Hilbert in Göttingen (to discuss with him mathematics). After graduating in chemistry in ETH in Zurich (1925) and receiving the doctorate in Budapest (1926) (his doctoral dissertation was devoted to the axiomatization of set theory – cf. below), he became *Privatdozent* at the University in Berlin (1927–1929), and next in Hamburg (1929–1930). In 1930 he left Germany and went to the USA.[1]

* The support of the Committee for Scientific Research (grant no. 1 H01A 042 27) is acknowledged.

[1] We are not describing further the life of von Neumann and stop at about 1930 because his disappointment with the investigations in the foundations of mathematic let to the fact that after 1930 he lost the interest in the foundational problems and turned his attention to other parts of mathematics, in particular to its applications (see Section 4). Note only that in 1930–1931 von Neumann was visiting lecturer at Princeton University in New Jersey, later a professor there. Since 1933 he was professor in the Institute for Advanced Study in Princeton. He died in 1957 at the age of 54.

ISBN 83-7431-015-4 ISSN 0860-150X

Talking about Hilbert's School we mean the group of mathematicians around Hilbert working in the foundations of mathematics and in the metamathematics (proof theory) in the frameworks of Hilbert's programme of justification of the classical mathematics (by finitistic methods).[2]

The main works of von Neumann from the period from 1922 (the data of his first publication) till 1931 concern mainly metamathematics as well as the quantum mechanics and the theory of operators (also Hilbert worked at that time in just those domains). In this paper we shall be interested in the former, i.e., works devoted to and connected with Hilbert's metamathematical programme.

Recall (to make clearer further considerations) that one of the steps in the realization of Hilbert's programme was the formalization (and in particular the axiomatization) of the classical mathematics (this was necessary for further investigations of mathematical theories by finitistic methods of the proof theory).

Main achievements of von Neumann connected with the ideology of Hilbert's School are the following:

- axiomatization of set theory and (connected with that) elegant theory of the ordinal and cardinal numbers as well as the first strict formulation of the principles of definitions by the transfinite induction,
- the proof (by finitistic methods) of the consistency of a fragment of the arithmetic of natural numbers,
- the discovery and the proof of the second incompleteness theorem (this was done independently of Gödel).

The rest of the paper will be devoted just to those items.

2. Foundations of set theory

Contribution of von Neumann devoted to the set theory consisted not only of having proposed a new elegant axiomatic system (extending the system of Zermelo-Fraekel ZFC) but also of having proposed several innovations enriching the system ZFC, in particular the definition of ordinal and cardinal numbers and the theory of definitions by transfinite induction.

The definition of ordinals and cardinals was given by von Neumann in the paper "Zur Einführung der transfiniten Zahlen" (1923) – it was his second publication. He has given there a definition of an ordinal number

[2] There is a rich literature on Hilbert's programme – see, e.g., (Murawski, 1999) and the literature indicated there.

which could "give unequivocal and concrete form to Cantor's notion of ordinal number" in the context of axiomatized set theories (cf. von Neumann, 1923). Von Neumann's ordinal numbers are – using the terminology of G. Cantor – representatives of order types of well ordered sets. In (1923) von Neumann wrote:

> What we really wish to do is to take as the basis of our considerations the proposition: 'Every ordinal is the type of the set of all ordinals that precede it.' But, in order to avoid the vague notion 'type', we express it in the form: 'Every ordinal is the set of the ordinals that precede it.' This is not a proposition proved about ordinals; rather, it would be a definition of them if transfinite induction had already been established.[3]

In this way one obtains the sequence $\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \ldots$ – i.e., von Neumann's ordinal numbers.

Those sets – as representatives – are in fact very useful, especially in the axiomatic set theory because they can be easily defined in terms of the relation $\in$ only and they are well order by the relation $\in$. They enable also an elegant definition of cardinal numbers. In the paper (1928) one finds the following definition: a well ordered set M is said to be an ordinal (number) if and only if for all $x \in M$, x is equal to the initial segment of M determined by x itself (as von Neumann wrote: $x = A(x; M)$). Elements of ordinal numbers are also ordinal numbers. An ordinal number is said to be a cardinal number if and only if it is not equipollent to any of its own elements.

In the paper (1923) von Neumann presupposed the notions of a well ordered set and of the similarity and then proved that for any well ordered set there exists a unique ordinal number corresponding to it. All that was done in a naïve set theory but a remark was added that it can be done also in an axiomatic set theory. And in fact von Neumann did it in papers (1928) and (1928a). To be able to do this in a formal way one needs the Axiom of Replacement (in the paper (1923) von Neumann called it Fraenkel's axiom). Since that time von Neumann was an staunch advocate of this axiom.

[3] Wir wollen eigentlich den Satz: „Jede Ordnungszahl ist der Typus der Menge aller ihr vorangehenden Ordnungszahlen" zur Grundlage unserer Überlegungen machen. Damit aber der vage Begriff „Typus" vermieden werde, in dieser Form: „Jede Ordnungszahl ist die Menge der ihr vorangehenden Ordnungszahlen." Dies ist kein bewiesener Satz über Ordnungszahlen, es wäre vielmehr, wenn die tranfinite Induktion schon begründet wäre, eine Definition derselben.

The problem of definitions by transfinite induction was considered by von Neumann in the paper "Über die Definition durch transfinite Induction, und verwandte Fragen der allgemeinen Mengenlehre" (1928). He showed there that one can always use definitions by induction on ordinal numbers and that such definitions are unequivocal. He proved that for any given condition $\varphi(x, y)$ there exists a unique function f whose domain consists of ordinals such that for any ordinal α one has $f(\alpha) = \varphi(F(f, \alpha), \alpha)$ where $F(f, \alpha)$ is a graph of the function f for arguments being elements of α.

Why is the discussed paper so important? For many years, in fact since the axiomatization of set theory by Zermelo, there were no formal counterparts of ordinal and cardinal numbers and this was the reason of avoiding them in the axiomatic set theory. It became a custom to look for ways of avoiding transfinite numbers and transfinite induction in mathematical reasonings (cf., e.g., Kuratowski, 1921 and 1922). Von Neumann's paper introduced a new paradigm which works till today. The leading idea of the paper was the will to give the set theory as wide field as possible.

The most important and known contribution of John von Neumann is undoubtedly a new approach and new axiomatization of the set theory. The main ideas connected with that appeared by von Neumann already in 1923 (he was then 23!!!). He described them in a letter to Ernst Zermelo from August 1923.[4] He wrote there that the impulse to his ideas came from a work by Zermelo "Untersuchungen über die Grundlagen der Mengenlehre. I" (1908) and added that in some points he went away from Zermelo's ideas, in particular

- the notion of 'definite property' had been avoided – instead the "acceptable schemas" for the construction of functions and sets had been presented,
- the axiom of replacement had been assumed – it was necessary for the theory of ordinal numbers (later von Neumann emphasized, like Fraenkel and Skolem, that it is needed in order to establish the whole series of cardinalities – cf. von Neumann 1928a),
- sets that are "too big" (for example the set of all sets) had been admitted but they were taken to be inadmissible as elements of sets (that sufficed to avoid the paradoxes).

About 1922–1923 while preparing a paper in which those ideas should be developed he contacted Abraham Fraenkel. The latter recalled this (already

[4] This letter was partly reproduced in Meschowski, 1967, 289–291.

after the death of von Neumann) in a letter to Stanisław Ulam in such a way:[5]

> Around 1922–23, being then professor at Marburg University, I received from Professor Erhard Schmidt, Berlin (on behalf of the *Redaktion* of the *Mathematische Zeitschrift*) a long manuscript of an author unknown to me, Johann von Neumann, with the title "Die Axiomatisierung der Mengenlehre", this being his eventual doctor[al] dissertation which appeared in the *Zeitschrift* only in 1928 (vol. 27). I was asked to express my views since it seemed incomprehensible. I don't maintain that I understood everything, but enough to see that this was an outstanding work and to recognize *ex ungue leonem*. While answering in this sense, I invited the young scholar to visit me (in Marburg) and discussed things with him, strongly advising him to prepare the ground for the understanding of so technical an essay by a more informal essay which should stress the new access to the problem and its fundament consequences. He wrote such an essay under the title "Eine Axiomatisierung der Mengenlehre" and I published it in 1925 in the *Journal für Mathematik* (vol. 154) of which I was then Associate Editor.

Before continuing the story let us explain that *ex ungue leonem* – spotting a lion from the claw – is an expression used by Daniel Bernoulli while talking about Newton two and a half centuries ago. Bernoulli was namely sent a mathematical paper without a name of the author but he immediately recognized that it has been written just by Newton.

Von Neuman began the paper "Eine Axiomatisierung der Mengenlehre" (1925) by writing:

> The aim of the present work is to give a logically unobjectionable axiomatic treatment of set theory. I would like to say something first about difficulties which make such an axiomatization of set theory desirable.[6]

He stressed explicitly three points mentioned in the letter to Zermelo.

The characteristic feature of the system of set theory proposed by von Neumann is the distinction between classes, "domains" (*Bereiche*) and sets (*Mengen*). Classes are introduced by the Principle of Comprehension – von Neumann seems to have regarded this principle as the quintessence of what he called "naïve set theory" (cf. von Neumann 1923, 1928, 1929). His approach to set theory was strongly based on the idea of limitation of size

[5] Letter from Fraenkel to Ulam in (Ulam, 1958).

[6] Das Ziel der vorliegenden Arbeit ist, eine logisch einwandfreie axiomtische Darstellung der Mengenlehre zu geben. Ich möchte dabei einleitend einiges über die Schwierigkeiten sagen, die einen derartigen Aufbau der Mengenlehre erwünscht gemacht haben.

according to which: a class is a set if and only if it is not "too big". The latter notion was described by the following axiom:

(*) *A class is "too big"* (in the terminology of (Gödel, 1940) – *is a proper class) if and only if it is equivalent to the class of all things.*

Hence a class of the cardinality smaller than the cardinality of the class of all sets is a set. Von Neumann states further that the above principle implies both the Axiom of Separation and the Axiom of Replacement. It implies also the well ordering theorem (he indicated it already in the letter to Zermelo). Indeed, according to the reasoning used in the Burali-Forti paradox, the class On of all ordinal numbers is not a set, hence by the above principle it is equipollent with the class V of all sets. In this way one obtains a strengthened version of the well ordering theorem, namely:

The class V of all sets can be well ordered.

In the paper "Die Axiomatisierung der Mengenlehre" (1928a) [this was in fact a "mathematical" version of the system of set theory announced in the paper (1925)] von Neumann observed that the principle gives also a global choice function F such that for any nonempty set A it holds: $F(A) \in A$.

The Axiom of Choice, being a consequence of (*), enabled von Neumann to introduce ordinal and cardinal numbers without the necessity of introducing any new primitive notions. It was in fact a realization of the idea he wrote about in 1923. He used here the Fraenkel's Axiom of Replacement.

Observe that the distinction between classes and sets appeared already by Georg Cantor – he wanted to eliminate in this way the paradox of the set of all sets. Cantor used to call classes "absolutely infinite multiplicities". But he gave no precise criterion of distinguishing classes and sets – it was given only by von Neumann. The latter has also shown that Cantor was mistaken when he claimed that the absolutely infinite multiplicities (e.g., the multiplicity of all ordinal numbers) cannot be treated as consistent objects.

In the original formulation of the set theory by von Neumann there are no notions of a set and a class. Instead, one has there the primitive notion of a function (and of the relation $\in$). Von Neumann claimed that it is in fact only a technical matter – indeed, the notions of a set and of a function are mutually definable, i.e., a set can be treated as a function with values 0 and 1 (the characteristic function of the set) and, vice versa, a function can be defined as a set of ordered pairs.

Add that in the von Neumann's system of the set theory there are no urelements – there are only pure sets and classes. On the other hand, among axioms there is the Axiom of Foundation introduced by Dimitri Mirimanoff in (1917).[7] This axiom guarantees that there are no infinite decreasing $\in$-sequences, i.e., such sequences that $\ldots \in x_n \in \ldots \in x_1 \in x_0$ and that there are neither sets x such that $x \in x$ nor sets x and y such that $x \in y$ and $y \in x$. This axiom implies that the system of set theory containing it becomes similar to the theory of types: one can say that the system ZF with the Axiom of Foundation can be treated as an extension of the (cumulative) theory of types to the transfinite types described in a simpler language than it was the case by Russell.

It should be noticed that von Neumann was one of the first who investigated metatheoretical properties of the axiomatic set theory. In particular he studied his own system from the point of view of the categoricity (1925) and of the relative consistency (1929). Probably he was also the first author who called attention to the Skolem paradox. According to von Neumann, this paradox stamps an axiomatic set theory "with the mark of unreality" and gives reasons to "entertain reservations" about it (cf. 1925).

Von Neumann wrote about the proof of the relative consistency of his system of the set theory in the paper "Über eine Widerspruchsfreiheitsfrage der axiomatischen Mengenlehre" (1929). He saw main difficulties in the axiom (*). Therefore he considered two axiomatic systems: system S which was his original system (hence with the axiom (*)) and the system S* which was von Neumann's system but with the Axiom of Replacement and the Axiom of Choice instead of the axiom (*). In the paper (1929) he proved that:

1. S* will remain consistent if one adds the Axiom of Foundation and does not admit urlements,
2. S* is a subsystem of such a system.

Hence von Neumenn proved the relative consistency of the Axiom of Foundation with respect to the system S*. It was in fact the first significant metatheoretical result on the set theory.

It is worth saying that in (1929) von Neumann developed the cumulative hierarchy in technical details. Using the Axiom of Foundation and the ordinal numbers he showed that the universe of sets can be divided into "levels" indexed by ordinal numbers. He introduced the notion of a rank of a set: a rank of a set x is the smallest ordinal number α such that the set x appears at the level α. This hierarchy is cumulative, i.e., lower levels are included in higher ones. The hierarchy can be precisely defined as follows:

[7] In fact it was for the first time discussed by Mirimanoff and Skolem and it was von Neumann who as the first formulated it explicitely.

$$V_0 = \emptyset,$$
$$V_{\alpha+1} = V_\alpha \cup \mathcal{P}(V_\alpha),$$
$$V_\lambda = \bigcup_{\alpha<\lambda} V_\alpha \quad \text{for } \lambda \in \lim,$$
$$V = \bigcup_{\alpha \in On} V_\alpha,$$
$$\text{rank}(x) = \mu\alpha(x \in V_\alpha).$$

It is worth adding here that von Neumann treated the Axiom of Foundation rather as a tool in the metatheoretical investigations of the set theory.

We are talking all the time about axioms of the set theory but no axioms have been given so far. It is time to do it!

Let us start by stating that the main idea underlying von Neumann's system of the set theory has been accepted with enthusiasm – in fact, it provided a remedium to too drastic restrictions put on objects of the set theory by the system ZF of Zermelo-Fraenkel (one was convinced that such strong restrictions are not needed in order to eliminate paradoxes; on the other hand, the restrictions put by ZF made the development of mathematics within ZF very difficult and unnatural). Nevertheless the system of von Neumann was not very popular among specialists – the reason was the fact that it was rather counter-intuitive and was based on a rather difficult notions (recall that the primitive notion of a function instead of the notion of a set was used there). Hence the need of reformulating the original system. That has been done by Paul Bernays: in (1937) he announced the foundations, and in a series of papers published in the period 1937–1958 (cf. 1937, 1941, 1942, 1958) he gave an extensive axiomatic system of the set theory which realized the ideas of von Neumann and simultaneously he suceeded to formulate his system in a language close to the language of the system ZF.

In (1937) he wrote:

> The purpose of modifying the von Neumann system is to remain nearer to the structure of the original Zermelo system and to utilize at the same time some of the set-theoretic concepts of the Schröder logic and of *Principia Mathematica* which have become familiar to logicians. As will be seen, a considerable simplification results from this arrangement.

The universe of the set theory consists by Bernays of two parts:

- sets denoted by $x, y, z, \ldots$,
- classes denoted by $A, B, C, \ldots$

Hence it is not an elementary system! There are two primitive notions: $\in$ (= to be an element of (to belong to) a set) and η (= to be an element (to belong to) a class). Hence one has two types of atomic formulas: $x \in y$ and $x\eta A$. There are also two groups of axioms: axioms about sets (they are analogous to axioms of Zermelo) and axioms characterizing classes. The very important feature of Bernays' axioms is the fact that there are only finitely many axioms and there are no axiom schemes.

In the work devoted to the consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis K. Gödel gave an axiomatic system of set theory which is in fact a modification of Bernays' system. Its main advantage is that it is an elementary system (i.e., it contains only one type of variables).[8]

Let us describe now in details the system NBG of Gödel. It is based on the idea that the variables vary over classes. Among classes we distinguish those classes that are elements of other classes. They are called sets and their totality is denoted by V. The remaining classes are called proper classes.

Define the class V as follows

$$x \in V \longleftrightarrow (\exists y)(x \in y).$$

Hence x is a set if and only if there exists a class y such that $x \in y$. Define also a notion of a function in the following way:

$$\text{Func}(r) \longleftrightarrow \forall x \forall y \forall z[(x, y) \in r \wedge (x, z) \in r \longrightarrow y = z].$$

The system NBG is based on the following nonlogical axioms:

- (Extensionality)
$$\forall x\, \forall y\, [\forall z(z \in x \longleftrightarrow z \in y) \longrightarrow x = y],$$
- (Axiom of Classes)
$$\exists x. \forall y\, [\exists z(y \in z) \longrightarrow y \in x],$$
- (Axiom of the Empty Set)
$$\exists x\, [\forall y\, (y \notin x) \wedge \exists z\, (x \in z)],$$
- (Pairing Axiom)
$$\forall x \in V\, \forall y \in V\, \exists z \in V\, [\forall u(u \in z \longleftrightarrow u = x \vee u = y)],$$

[8] It is worth noting here that the idea of using in the system of von Neumann – Bernays only one type of variables and one membership relation is due to Alfred Tarski – cf. (Mostowski, 1939, p. 208) and (Mostowski, 1949, p. 144).

- (Axiom Scheme of Class Existence) if Φ is a formula with free variables $v_1, \dots, v_n$, then the following formula
$$\forall v_1, \dots, v_n \in V\, \exists z\, \forall x\, [x \in z \longleftrightarrow (x \in V \wedge \Phi^{(V)}(x, v_1, v_2, \dots, v_n))]$$
is an axiom (note that one cannot quantifier in Φ over class variables!): $\Phi^{(V)}$ denotes the relativization of Φ to the class V,
- (Axiom of Union)
$$\forall x \in V\, \exists y \in V\, \forall u\, [u \in y \longleftrightarrow \exists v (u \in v \wedge v \in x)],$$
- (Power Set Axiom)
$$\forall x \in V\, \exists y \in V\, \forall u\, (u \in y \longleftrightarrow u \subseteq x),$$
- (Infinity Axiom)
$$\exists x \in V\, [\emptyset \in x \wedge \forall u \in x\, \forall v \in x\, (u \cup \{v\} \in x)],$$
- (Axiom of Replacement)
$$\forall x \in V\, \forall r\, [Func(r) \longrightarrow \exists y \in V\, \forall u\, (u \in y \longleftrightarrow \exists v \in x\, ((v, u) \in r)],$$
- (Axiom of Foundation)
$$\forall x [x \neq \emptyset \longrightarrow \exists y \in x\, (x \cap y = \emptyset)].$$

We one adds to this system the following Axiom of Global Choice (in a strong version):

$$\exists x [Func(x) \wedge \forall y \in V [y \neq \emptyset \longrightarrow \exists z (z \in y \wedge (y, z) \in x)]]$$

then one obtains the system denoted as NBGC.

It has turned out that the axiom scheme of class existence can be replaced by the following (finitely many!) axioms:

$$\exists a\, \forall x, y \in V [(x, y) \in a \longleftrightarrow x \in y]$$

(it says that a is a graph of the membership relation $\in$ for sets),

$$\forall a\, \forall b\, \exists c\, \forall x [x \in c \longleftrightarrow (x \in a \wedge x \in b)]$$

(it defines the intersection of classes),

$$\forall a\, \exists b\, \forall x \in V\, [x \in b \longleftrightarrow x \notin a]$$

(it defines the complement of a class),

$$\forall a\, \exists b\, \forall x \in V\, [x \in b \longleftrightarrow \exists y \in V ((x, y) \in a)]$$

(it defines the left domain of a relation),

$$\forall a\, \exists b\, \forall x, y \in V\, [(x, y) \in b \longleftrightarrow x \in a),$$
$$\forall a\, \exists b\, \forall x, y, z \in V\, [(x, y, z) \in b \longleftrightarrow (y, z, x) \in a],$$
$$\forall a\, \exists b\, \forall x, y, z \in V\, [(x, y, z) \in b \longleftrightarrow (x, z, y) \in a].$$

The systems NBG and NBGC have very nice metamathematical properties, in particular:

- NBG (NBGC) is finitely axiomatizable (observe that the Zermelo-Fraenkel system ZF is not finitely axiomatizable!),
- NBG is a conservative extension of ZF with respect to formulas saying about sets, i.e., for any formula φ of the language of set theory:
$$\mathrm{ZF} \vdash \varphi \qquad \text{if and only if} \qquad \mathrm{NBG} \vdash \varphi^{(V)}$$

where $\varphi^{(V)}$ denotes the relativization of φ to the class V of all sets (and similarly for NBGC and ZFC where the latter symbol denotes the theory ZF plus the Axiom of Choice AC),

- NBG is consistent if and only if ZF is consistent (and similarly for NBGC and ZFC).

3. Consistency proof for arithmetic

One of the main aims of Hilbert's programme was the consistency proof (by save finitary methods) for the whole classical mathematics. Students of Hilbert took this task and soon the first partial results appeared. The first work in this direction was the paper by Wilhelm Ackermann (1924) where he gave a finitistic proof of the consistency of arithmetic of natural numbers without the axiom (scheme) of induction.[9]

The next attempt to solve the problem of the consistency was the paper "Zur Hilbertschen Beweistheorie" (1927) by von Neumann. He used another formalism than that in (Ackermann, 1924) and, similarly as Ackermann, proved in fact the consistency of a fragment of arithmetic of natural numbers obtained by putting some restrictions on the induction. We cannot consider here the (complicated) technical details of von Neumann's proof. It is worth mentioning that in the introductory section of von Neumann's (1927) a nice and precise formulation of aims and methods of Hilbert's proof

[9] In fact it was a much weaker system than the usual system of arithmetic but the paper provided the first attempt to solve the problem of consistency. Later in the paper (1940) Ackermann proved the consistency of the full arithmetic of natural numbers by using methods from his paper (1924) and the transfinite induction.

theory was given. It indicated how was at that time the state of affairs and how Hilbert's programme was understood. Therefore we shall quote the appropriate passages.

Von Neumann writes that the essential tasks of the proof theory are (cf. von Neumann, 1927, 256–257):

> I. First of all, one wants to give a proof of the consistency of the classical mathematics. Under 'classical mathematics' one means the mathematics in the sense in which it was understood before the begining of the criticism of the set theory. All settheoretic methods essentially belong to it but not the proper abstract set theory. [...]
> II. By the end the whole language and proving machinary of the classical mathematics should be formalized in an absolutely strong way. The formalism cannot be too narrow.
> III. Then one must prove the consistency of this system, i.e., one should show that certain formulas of the formalism just described can never be "proved".
> IV. One should always strongly distinguish here between various types of "proving": between formal ("mathematical") proving in a given formal system and contents ("metamathematical") proving [of statements] about the system. Whereas the former one is an arbitrarily defined logical game (which should to a large extent be analogues to the classical mathematics), the latter is a chain of directly evident contents insights. Hence this "contents proving" must proceed according to the intuitionistic logic of Brouwer and Weyl. The proof theory should so to speak construct classical mathematics on the intuitionistic base and in this way lead the strict intuitionism ad absurdum.[10]

[10] I. In erster Linie wird der Nachweis der Widerspruchsfreiheit der klassischen Mathematik angestrebt. Unter „klassischer Mathematik" wird dabei die Mathematik in demjenigen Sinne verstanden, wie sie bis zum Auftreten der Kritiker der Mengenlehre anerkannt war. Alle mengentheoretischen Methoden gehören im wesentlichen zu ihr, nicht aber die eigentliche abstrakte Mengenlehre. [...]

II. Zu diesem Zwecke muß der ganze Aussagen- und Beweisapparat der klassischen Mathematik absolut streng formalisiert werden. Der Formalismus darf keinesfalls zu eng sein.

III. Sodann muß die Widerspruchsfreiheit dieses Systems nachgewiesen werden, d.h. es muß gezeigt werden, daß gewisse Aussagen „Formeln" innerhalb des beschriebenen Formalismus niemals „bewiesen" werden können.

IV. Hierbei muß stets scharf zwischen verschiedenen Arten des „Beweisens" unterschieden werden: Dem formalistischen („mathematischen") Beweisen innerhalb des formalen Systems, und dem inhaltlichen („metamathematischen") Beweisen über das System. Während das erstere ein willkürlich definiertes logisches Spiel ist (das freilich mit der klassischen Mathematik weitgehend analog sein muß), ist das letztere eine Verkettung unmittelbar evidenter inhaltlicher Einsichten. Dieses „inhaltliche Beweisen" muß also ganz im Sinne der Brouwer-Weylschen intuitionistischen Logik verlaufen: Die Beweistheorie soll sozusagen auf intuitionistischer Basis die klassische Mathematik aufbauen und den strikten Intuitionismus so ad absurdum führen.

Note that von Neumann identifies here finitistic methods with intuitionistic ones. This was then current among members of the Hilbert's school. The distinction between those two notions was to be made explicit a few years later – cf. (Hilbert and Bernays, 1934, pp. 34 and 43) and (Bernays 1934, 1935, 1941a), see also (Murawski, 2001).

As an interesting detail let us add that in response to the paper (1927) by von Neumann it was reacted critically by Stanisław Leśniewski who published the paper "Grundzüge eines neuen Systems der Grundlagen der Mathematik" (1929) in which he critically analyzed various attempts to formalize logic and mathematics. Leśniewski among others expresses there his doubts concerning the meaning and significance of von Neumann's proof of the consistency of (a fragment of) arithmetic and constructs – to maintain his thesis – a "counterexample", namely he deduce (on the basis of von Neumann's system) two formulas a and $\neg a$, hence an inconsistency.

Von Neumann answered to Leśniewski's objections in the paper "Bemerkungen zu den Ausführungen von Herrn St. Leśniewski über meine Arbeit 'Zur Hilbertschen Beweistheorie'" (1931). Analyzing the objections of Leśniewski he came to the conclusion that there is in fact a misunderstanding resulting from various ways in which they both understand principles of formalization. He used also the occasion to fulfill the gap in his paper (1927).

Add also that looking for a proof of the consistency of the classical mathematics and being (still) convinced of the possibility of finding such a proof (in particular a proof of the consistency of the theory of real numbers) von Neumann doubted whether there are any chances to find such a proof for the set theory – cf. his paper (1929).

4. Von Neumann and Gödel's second incompleteness theorem

How much von Neumann was engaged in the realization of Hilbert's programme and how high was his position in this group can be judged from the fact that just he has been invited by the organizers of the Second Conference on the Epistemology of Exact Sciences (organized by Die Gesellschaft für Empirische Philosophie)[11] held in Königsberg, 5–7th September 1930, to give a lecture presenting formalism – one of the three main trends in the con-

[11] This conference was organized together with the 91st Convention of the Society of German Scientists and Physicians (Gesellschaft deutscher Naturforscher und Ärzte) and the 6th Conference of German Mathematicians and Physicists (Deutsche Physiker- und Mathematikertagung).

temporary philosophy of mathematics and the foundations of mathematics founded by Hilbert. The other two main trends: logicism and intuitionism were presented by Rudolf Carnap and Arend Heyting, resp.

In his lecture "Die formalistische Grundlegung der Mathematik" (cf. 1931a) von Neumann recalled basic presuppositions of Hilbert's programme and claimed that thanks to the works of Russell and his school a significant part of the tasks put by Hilbert has already been realized. In fact the unique task that should be fulfilled now is "to find a finitistically combinatorial proof of the consistency of the classical mathematics". And he added that this task turned out to be difficult. On the other hand, partial results obtained so far by W. Ackermann, H. Weyl and himself make possible to cherish hopes that it can be realized. He finished his lecture by saying: "Whether this can be done for a more difficult and more important system of [the whole] classical mathematics will show the future."

On the last day of the conference, i.e., on 7th September 1930, a young Austrian mathematician Kurt Gödel announced his recent (not yet published) results on the incompleteness of the system of arithmetic of natural numbers and richer systems.

It seems that the only participant of the conference in Königsberg who immediately grasped the meaning of Gödel's theorem and understood it was von Neumann. After Gödel's talk he had a long discussion with him and asked him about details of the proof. Soon after coming back from the conference to Berlin he wrote a letter to Gödel (on 20th November 1930) in which he announced that he had received a remarkable corollary from Gödel's First Theorem, namely a theorem on the unprovability of the consistency of arithmetic in arithmetic itself. In the meantime Gödel developed his Second Incompleteness Theorem and included it in his paper "Über formal unentscheidbare Sätze der 'Principia Mathematica' und verwandter Systeme. I" (cf. Gödel, 1931). In this situation von Neumann decided to leave the priority of the discovery to Gödel.

5. Concluding remarks

Gödel's incompleteness results had great influence on von Neumann's views towards the perspectives of investigations on the foundations of mathematics. He claimed that "Gödel's result has shown the unrealizability of Hilbert's program" and that "there is no more reason to reject intuitionism" (cf. his leter to Carnap of 6th June 1931 – see Mancosu, 1999, 39–41). He added in this letter:

> Therefore I consider the state of the foundational discussion in Königsberg to be outdated, for Gödel's fundamental discoveries have brought the question to a completely different level. (I know that Gödel is much more careful in the evoluation of his results, but in my opinion on this point he does not see the connections correctly).

Incompleteness results of Gödel changed the opinions cherished by von Neumann and convinced him that the programme of Hilbert cannot be realized. In the paper "The Mathematician" (1947) he wrote:

> My personal opinion, which is shared by many others, is, that Gödel has shown that Hilbert's program is essentially hopeless.

Another reason for the disappointment of von Neumann with the investigations in the foundations of mathematics could be the fact that he became aware of the lack of categoricity of set theory, i.e., that there exist various nonisomorphic models of the set theory. The latter fact implies that it is impossible to describe the world of mathematics in a unique way. In fact there is no absolute description, all descriptions are relative.

Not only von Neumman was aware of this feature of the set theory. Also Fraenkel and Thoralf Skolem realized this. And they have proposed various measures. In particular Fraenkel in his very first article "Über die Zermelosche Begründung der Mengenlehre" (1921) sought to render set theory categorical by introducing his Axiom of Restriction, inverse to the completeness axiom that Hilbert had proposed for geometry in 1899. Whereas Hilbert had postulated the existence of a maximal model satisfying his other axioms, Fraenkel's Axiom of Restriction asserted that the only sets to exist were those whose existence was implied by Zermelo's axioms and by the Axiom of Replacement. In particular, there were no urelements. One should add that Fraenkel did not distinguish properly a language and a metalanguage and confused them.

The approach of Skolem was different – but we will not go into technical details here.[12]

Von Neumann also examined the possible categoricity of set theory. In order to render it as likely as possible that his own system was categorical, he went beyond Mirimanoff and augmented it by the axiom stating that there are no infinite descending $\in$-sequences. He recognized that his system would surely lack categoricity unless he excluded weakly inaccessible cardinals (i.e., regular cardinals with an index being a limit ordinal). Von Neumann

[12] See, e.g., Moore, 1982, Section 4.9.

rejected also the Fraenkel's Axiom of Restriction as untenable because it relied on the concept of subdomain and hence on inconsistent "naïve" set theory. He was also aware of the difficulties implied by Löwenheim-Skolem theorem.

Von Neumann treated the lack of categoricity of set theory, certain relativism of it as an argument in favor of intuitionism (cf. his 1925). He stressed also the distance between the naïve and the formalized set theory and called attention to the arbitrariness of restrictions introduced in the axiomatic set theory (cf. 1925, 1928a, 1929). He saw also no rescue and no hope in Hilbert's programme and his proof theory – in fact the latter was concerned with consistency and not with categoricity.

One should notice here that von Neumann's analyses lacked a clear understanding of the difference and divergence between first-order and second-order logic and their effects on categoricity. Today it is known, e.g., that Hilbert's axioms for Euclidian geometry and for the real numbers as well as Dedekind-Peano axioms for the arithmetic of natural numbers are categorical in second-order logic and non-categoric in the first-order logic. Only Zermelo (perhaps under the influence of Hilbert[13]) claimed that the first-order logic is insufficent for mathematics, and in particular for the set theory. It became the dominant element in Zermelo's publications from the period 1929–1935. It is worth noting here that he spoke about this for the first time in his lectures held in Warsaw in May and June 1929.

After 1931 von Neumann ceased publishing on the mathematical logic and the foundations of mathematics - he came to the conclusion that a mathematician should devote his attention to problems connected with the applications. In (1947) he wrote:

> As a mathematical discipline travels far from its empirical source, or still more, if it is a second and third generation only indirectly inspired by ideas coming from "reality", it is beset with very grave dangers. It becomes more and more purely aestheticizing, more and more purely *l'art pour l'art*. ... In other words, at a great distance from its empirical source, or after much "abstract" inbreeding, a mathematical subject is in danger of degeneration.

[13] Hilbert and Ackermann wrote in (1928): "As soon as the object of investigation becomes the foundation of ... mathematical theories, as soon as we went to determine in what relation the theory stands to logic and to what extent it can be obtained from purely logical operations and concepts, then second-order logic is essential." In particular they defined the set-theoretic concept of well-ordering by means of second-order, rather than first-order, logic.

References

Ackermann W., 1924, Begründung des „tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit, *Mathematische Annalen* 93, 1–36.

Ackermann, W., 1940, Zur Widerspruchsfreiheit der Zahlentheorie, *Mathematische Annalen* 117, 162–194.

Bernays, P., 1934, 'Quelques points essentiels de la métamathématiques', *L'Enseignement mathématique* 34, 70–95.

Bernays, P., 1935, Hilberts Untersuchungen über die Grundlagen der Arithmetik, in: Hilbert, D. *Gesammelte Abhandlungen*, Verlag von Julius Springer, Berlin, Bd. 3, pp. 196–216.

Bernays P., 1937, A system of axiomatic set theory. I, *Journal of Symbolic Logic* 2, 65–77.

Bernays P., 1941, A system of axiomatic set theory. II, *Journal of Symbolic Logic* 6, 1–17.

Bernays, P., 1941a, Sur les questions méthodologiques actuelles de la théorie hilbertienne de la démonstration, in: Gonseth, F. (Ed.) *Les entretiens de Zurich sur les fondements et la méthode des sciences mathématiques, 6–9 décembre 1938*, Leemann, Zurich 1941, pp. 144–152; Discussion, pp. 153–161.

Bernays P., 1942, A system of axiomatic set theory. III, *Journal of Symbolic Logic* 7, 65–89, 133–145.

Bernays P., 1958, *Axiomatic Set Theory*, North-Holland Publishing Company, Amsterdam.

Fraenkel A., 1921, Über die Zermelosche Begründung der Mengenlehre, *Jahresbericht der Deutschen Mathematiker-Vereinigung* (Angelegenheiten) 30, 97–98.

Gödel, K., 1931, 'Über formal unentscheidbare Sätze der 'Principia Mathematica' und verwandter Systeme. I', *Monatshefte für Mathematik und Physik* 38, 173–198. Reprinted with English translation 'On Formally Undecidable Propositions of Principia Mathematica and Related Systems' in: Gödel, K. *Collected Works*, vol. I, ed. by Feferman, S. *et al.*, Oxford University Press, New York and Clarendon Press, Oxford, 1986, pp. 144–195.

Gödel K., 1940, *The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis with the Axioms of Set Theory*, Princeton University Press, Princeton, N.J.

Hilbert, D. and Ackermann, W., 1928, *Grundzüge der theoretischen Logik*, Springer Verlag, Berlin. English translation: *Principles of Mathematical Logic*, AMS Chelsea Publishing, American Mathematical Society, Providence, Rhode Island 1958.

Hilbert, D. and Bernays, P., 1934–1939, *Grundlagen der Mathematik*, Bd. I 1934, Bd. II 1939, Springer-Verlag, Berlin.

Kuratowski K., 1922, Sur la notion de l'ordre dans la théorie des ensembles, *Fundamenta Mathematicae* 2, 161–171.

Kuratowski K., 1923, Une méthode d'elimination des nombres transfinis des raisonnements mathématiques, *Fundamenta Mathematicae* 3, 76–108.

Leśniewski S., 1929, Grundzüge eines neuen Systems der Grundlagen der Mathematik, *Fundamenta Mathematicae* 14, 1–81.

Mancosu, P., 1999, Between Vienna and Berlin: The Immediate Reception of Gödel's Incompleteness Theorems, *History and Philosophy of Logic* 20, 33–45.

Meschkowski H., 1967, *Probleme des Unendlichen. Werk und Leben Georg Cantors*, Vieweg, Braunschweig. Reproduced with additions as *Georg Cantor. Leben, Werk und Wirkung*, Bibliographisches Institut, Mannheim 1983.

Mirimanoff D., 1917, Les antinomies de Russell et de Burali-Forti et le probléme fondamental de la théorie des ensembles, *L'Enseignement Mathématique* 19, 37–52.

Moore G. H., 1982, *Zermelo's Axiom of Choice. Its Origins, Development, and Influence*, Springer Verlag, New York/Heidelberg/Berlin.

Mostowski A., 1939, Über die Unabhängighkeit des Wohlordnungssatzes vom Ordnungsprinzip, *Fundamenta Mathematicae* 32, 201–252.

Mostowski A., 1949, An undecidable arithmetical statement, *Fundamenta Mathematicae* 36, 143–164.

Murawski R., 1999, *Recursive Functions and Metamathematics*, Kluwer Academic Publishers, Dordrecht/Boston/London 1999.

Murawski R., 2001, On proof of the consistency of arithmetic, *Studies in Logic, Grammar and Rhetoric* 4 (17), 41–50.

von Neumann J., 1923, Zur Einführung der transfiniten Zahlen, *Acta Szeged* 1, 199–208. Reprinted in: von Neumann J., 1961, 24–33.

von Neumann J., 1925, Eine Axiomatisierung der Mengenlehre, *Journal für Mathematik* 154, 219–240. Reprinted in: von Neumann J., 1961, 34–57.

von Neumann J., 1927, Zur Hilbertschen Beweistheorie, *Mathematische Zeitschrift* 26, 1–46. Reprinted in: von Neumann J., 1961, 256–302.

von Neumann J., 1928, Über die Definitionen durch transfinite Induktion, und verwandte Fragen der allgemeinen Mengenlehre, *Mathematische Annalen* 99, 373–391. Reprinted in: von Neumann J., 1961, 320–338.

von Neumann J., 1928a, Die Axiomatisierung der Mengenlehre, *Mathematische Zeitschrift* 27, 669–752. Reprinted in: von Neumann J., 1961, 339–422.

von Neumann J., 1929, Über eine Widerspruchsfreiheitsfrage der axiomatischen Mengenlehre, *Journal für Mathematik* 160, 227–241. Reprinted in: von Neumann J., 1961, 494–508.

von Neumann J., 1931, Bemerkungen zu den Ausführungen von Herrn St. Leśniewski über meine Arbeit "Zur Hilbertschen Beweistheorie", *Fundamenta Mathematicae* 17, 331–334. Reprinted in: von Neumann J., 1961a, 230–233.

von Neumann J., 1931a, Die formalistische Grundlegung der Mathematik, *Erkenntniss* 2, 116–121. Reprinted in: von Neumann J., 1961a, 234–239.

von Neumann J., 1947, The Mathematician, in: *The Works of the Mind*, ed. by R. B. Heywood, University of Chicago Press, Chicago, 180–196. Reprinted in: von Neumann J., 1961, 1–9.

von Neumann J., 1961, *Collected Works*, vol. I: *Logic, Theory of Sets and Quantum Mechanics*, Pergamon Press, Oxford/London/New York/Paris.

von Neumann J., 1961a, *Collected Works*, vol. II: *Operators, Ergodic Theory and Almost Periodic Functions in a Group*, Pergamon Press, Oxford/London/New York/Paris.

Ulam S., 1958, John von Neumann, 1903–1957, *Bulletin of the American Mathematical Society* 64 (May Supplement), 1–49.

Zermelo, E., 1908, Untersuchungen über die Grundlagen der Mengenlehre. I, *Mathematische Annalen* 65, 261–281.

Roman Murawski
Adam Mickiewicz University
Faculty of Mathematics and Comp. Sci.
ul. Umultowska 87
61-614 Poznań, Poland
e-mail: rmur@amu.edu.pl

Jerzy Mycka
University of Maria Curie-Sklodowska, Lublin

EMPIRICAL ASPECTS OF COMPUTABILITY THEORY
that is: does the Univers compute better than (thoughtless) man?

Abstract. The paper deals with the question how the shape of physical reality affects the classification of problems into computable and noncomputable. First A. Turing's idea of computability is recalled. On the basis of this theory a possibility of revising the notion of computability will be presented. The starting point will be a conception of a physical system processing information understood as parameters of the system elements. Appropriate for this approach models of computability which are not brought to Turing machines will be pointed out. As the major objection to the models of the above-mentioned type the Gandy thesis will be considered. The paper justifies methodologically as plausible the search of the problems exceeding, in the sense of computability, the Turing model.

Computability

The idea of mechanical solving of problems goes back to the distant past in science. We will not, however, deal with the historical outline of this issue, but immediately turn to the notions accompanying the examinations of computability which were coined by Alan Turing [18].

The model of computability called the Turing machine, described by means of an informal language, can be defined as follows: the machine comprises an infinite tape divided into identical cells, which is to store input data, output data and working information. All the elements on the tape are strings (sequences of symbols); at the same time there is a rule of placing one symbol into one cell. Without a loss of generality a particular alphabet is usually chosen as a range of symbols permissible to build strings. Practically the binary (zero, one) alphabet is often chosen, it allows a convenient and consistent representation of data. Moreover, the construction of the machine requires a description of a finite set of states from which an element, indicating a current situation (state) of the machine, originates.

ISBN 83-7431-015-4 ISSN 0860-150X

The Turing machine works in steps which are identical as to the time of lasting. At every step the machine reads the content of a current cell (pointed out by the head), then it changes the content according to a read symbol and a present state. The same information serves successively to change the state and to move the head to the next left or right square. Particular states are distinguished as final; if the machine is in one of them, it finishes its work. The classical model of the Turing machine requires the number of symbols on the tape (different from the empty symbol) to be finite at every moment of the machine work.

A problem is computable, according to A. Turing, if its solution can be found due to the properly constructed Turing machine. Let us notice that the whole process of computing described above can be easily imagined as the work of a man who, by means of a sheet of paper and a pencil, realizes (thoughtlessly) consecutive changes of symbols according to strict rules. Robert Soare [17] stressed the character of Turing's computability, naming the computing subject 'computor', to emphasize the idealization of human activity which was used here. An analysis of the machine potentials (physical systems) does not seem to be the Turing's aim. In agreement with the previous remarks the construction of Turing machines encloses rather the possibilities of 'an ideal mathematician' activity.

Turing machines defined in the following way allow a comprehensive analysis of computability problems. The most important are: distinguishing the issues that cannot be explained by means of the Turing machine (unsolvable problems) and pointing out time and spatial restrictions originating from the nature of a solved issue (theory of complexity).

Of course, the Turing machine is not the only one model of computable processes. Just to illustrate, there are the Markov algorithms [6], the Church λ-calculus [1] or partially recursive functions all of which were proposed independently. In the course of research for all the models the thesis has been verified that they are identical as far as the scope of computability is concerned (compare [10]). Using an informal language: every issue explained in one of these models can be solved in any other model as well.

Let us digress from the subject here. The models are equivalent as to the power but different as regards a means of expression. The Markov algorithms and the Church calculus are the models based on the description of particular operations on strings. Nevertheless, it is an approach which is very close to the one usually accepted by adherents of formalism, regarding the foundations of mathematics. Then the description of calculations reflecting human activities – the Turing machine – leads us into the areas that are close to mathematical intuitionism. The world of mathematical functions in the natural domain, yet, seems to be the domain of platonism. As it can be suggested by this short enumeration, a preference for using one computability model may be connected with philosophical orientation regarding the foundations of mathematics.

The results of researches concerning different computability models led to formulate the Church-Turing thesis [2]. It will be recalled at this stage in order to avoid misunderstandings. The above thesis claims that our informal notion of an effective computing is parallel to the precise Turing machine mathematical definition. The conditions usually associated with efficiency are connected with a finite number of activity instructions (rules) which are represented by finite strings accompanied by action (in a proper case) in a finite number of steps with a possibility (in general) of realizing the algorithm by a man without using intuition, creativity, or direct insight into the essence. Let us pay attention to what the Church-Turing thesis does not say: it does not claim that other types of computing are not possible. It just points out that the calculations that are effective are related to the above mentioned models.

Hypercomputabilty

In the light of the above mentioned facts the question arises, essential both theoretically and practically – is there a form of an ineffective calculation, yet possible to be realized? Ordinary procedures and models of this property are described by the notion of hypercomputability. Since an effectiveness of calculations seems to represent a human activity description by means of the theory of computability language, thus ineffective solutions are sought in the world of physics.

In agreement with these remarks we will present the possibilities of revising the notion of computability. The starting point will be a concept of physical (artificial or natural) system transferring the information perceived as parameters of elements of the system. A mechanical character of such an approach agrees with a traditional name 'computer'.

Let us begin from introducing examples of ineffective computability models. A straightforward modification of the Turing machine, called the accelerating Turing machine, will be presented as the first model. The description of its structure is identical to the basic Turing machine. The change concerns the timing pattern of realized steps. Every consecutive step is realized in the time equal to the half time of the preceding step. The machine working in the following way is able to produce an infinite number of steps during the first two initial units of time. This feature allows solving the problem of halting some Turing machine by the following widening of the set of its instructions: firstly, mark with 0 symbol the chosen cell of the machine; in case of introducing the final instruction – change the input of this cell into 1. After two units of time an examination of this cell allows us to state whether the Turing machine stopped after finite or infinite number of steps. Because the halting problem is not effectively computable, a new model spreads significantly the limits of the notion of computability.

Let us turn into the other field of nonefective computation: namely analog computation. The basic model in this field is Shannon's General Purpose Analog Computer [16].

The General Purpose Analog Computer (GPAC) is a computer whose computation evolves in continuous time. The outputs are generated from the inputs by means of a dependence defined by a finite directed graph (not necessarily acyclic) where each node is one of the following boxes.

- *Integrator*: a two-input, one-output unit with a setting for initial condition. If the inputs are unary functions u, v, then the output is the Riemann-Stieljes integral $\lambda t. \int_{t_0}^{t} u(x)dv(x) + a$, where a and t_0 are real constants defined by the initial settings of the integrator.
- *Constant multiplier*: a one-input, one-output unit associated to a real number. If u is the input of a constant multiplier associated to the real number k, then the output is ku.
- *Adder*: a two-input, one-output unit. If u and v are the inputs, then the output is $u + v$.
- *Multiplier*: a two-input, one-output unit. If u and v are the inputs, then the output is uv.
- *Constant function*: a zero-input, one-output unit. The value of the output is always 1.

The next important in this context model of analog computation is Rubel's Extended Analog Computer (EAC) [13, 14]. This model is similar to the GPAC, but it allows, in addition, other types of units, e.g. units that solve boundary value problems (here we allow several independent variables because Rubel is not seeking any equivalence with existing models). The EAC permits all the operations of ordinary analysis, except the unrestricted taking of limits. The new units add an extended computational power relatively to the GPAC. For example, the EAC can solve the Dirichlet problem for Laplace's equation in the disk and can generate the Γ function (it is known that the GPAC cannot solve these problems [13]).

The model which is similar in the analog realm to classical natural recursive functions is the system of real recursive functions. Here we present a version given in [9] and based on the work of C. Moore'a [7]. The below definition is based on vector operations.

The set of real recursive vectors is generated from the real recursive scalars $0, 1, -1$ and the real recursive projections $I_n^i(x_1, \ldots, x_n) = x_i, 1 \leq i \leq n, n > 0$, by the operators:

1. composition: if f is a real recursive vector with n k-ary components and g is a real recursive vector with k m-ary components, then the vector with n m-ary components $(1 \leq i \leq n)$

   $$\lambda x_1 \ldots x_m . f_i(g_1(x_1, \ldots, x_m), \ldots, g_k(x_1, \ldots, x_m))$$

   is real recursive.
2. differential recursion: if f is a real recursive vector with n k-ary components and g is a real recursive vector with n $k+n+1$-ary components, then the vector h of n $k+1$-ary components which is the solution of the Cauchy problem for $1 \leq i \leq n$

   $$h_i(x_1, \ldots, x_k, 0) = f_i(x_1, \ldots, x_k),$$

   $$\partial_y h_i(x_1, \ldots, x_k, y) = g_i(x_1, \ldots, x_k, y, h_1(x_1, \ldots, x_k, y), \ldots, h_n(x_1, \ldots, x_k, y))$$

   is real recursive whenever h and its derivative are continuous in y on the largest interval containing 0 in which a unique solution exists except for a countable set of isolated points of discontinuity (of its derivative) where only one analytical continuation exists.
3. infinite limits: if f is a real recursive vector with n $k+1$-ary components, then the vectors h, h', h'' with n k-ary components $(1 \leq i \leq n)$

   $$h_i\,(x_1, \ldots, x_k) = \lim_{y \to \infty} f_i(x_1, \ldots, x_k, y),$$

   $$h'_i\,(x_1, \ldots, x_k) = \liminf_{y \to \infty} f_i(x_1, \ldots, x_k, y),$$

   $$h''_i(x_1, \ldots, x_k) = \limsup_{y \to \infty} f_i(x_1, \ldots, x_k, y),$$

   are real recursive, whenever these limits are defined for all $1 \leq i \leq n$.[1]
4. Arbitrary real recursive vectors can be defined by assembling scalar real recursive components.

[1] These concepts are defined in the completion of the real numbers $R \cup \{-\infty, +\infty\}$.

5. If f is a real recursive vector, than each of its components is a real recursive scalar.

Let us discuss the definition carefully. For differential recursion we restrict a domain to an interval of continuity. This will preserve the analiticity of functions in the process of defining. Moreover, this operator gives the same class C^k for a defined function as the given functions come from. This eliminates a possibility of defining such functions as $\lambda x.|x|$.

Let us point out the fact that this definition has as its feature the property of a real recursive computable equation relation. It is not a general case for an analog computation.

From the physical point of view with such definition we are ready to use only a finite amount of energy. We excluded here the possibility of operations on undefined functions: our functions are strict in the meaning that for undefined arguments they are also undefined. But to obtain some interesting functions we should improve the power of our system by an addition of the operators of infinite limits. Let us point out that introducing infinite limits gets discontinuous functions.

Infinity versus computability

If the Turing machine seems to be practically realized in the physical world, the above-mentioned models raise considerable doubts in this respect. They can be described precisely by means of the Gandy thesis. It is usually formulated in the following way: everything that can be computed by discrete deterministic mechanical device, can be calculated by the Turing machine as well. Let us notice that, using the rule of contraposition in this statement, we obtain an equivalent formulation: a problem impossible to be solved by the Turing machine will not be computable by means of any discrete and deterministic device. As the models mentioned above exceed the limits of classical computability, we would have to, on the strength of the above thesis, admit their practical 'non-realizability' in deterministic and discrete world.

Before an attempt of estimating the Gandy thesis let us consider a phenomenon of solving by these models the issues noncomputable in Turing's sense. An accelerating Turing machine is the simplest case for an analysis. Obtaining the message about a finite or infinite number of operations conducted by the machine is not connected with any refined procedure of examining this property. The machine just realizes its functions in a limited or unlimited number of steps. Appearing of results is a consequence of its

ability to realize an infinite number of steps in finite time. The power of this model lies in achieving a result due to find a limit in restricted time. The same possibilities appear for real recursive functions and EAC as a result of inscribing them into a construction of infinite limits operators. It is worth noticing that the same reason causes the mentioned models to be ineffective, namely they do not fulfill the requirement of a finite number of computing steps.

However, recognizing principles of the models constructions and the Gandy thesis to be mutually excluding, seems to be premature. Although discretion and determinism of computing are guaranteed in the thesis, an idea of an infinite number steps device is not rejected explicitly. Assuming implicitly such a limitation is connected with a conviction that an infinite number of steps requires infinite time. However, this statement is not at all obvious and is more connected with qualities of the physical world that surrounds us than with an inner structure of the proposed models.

Infinity versus physics

In the light of previous considerations, for establishing the boundaries of practically realized models of computability, the nature of the material world becomes essential. It is important, however, to become aware of an obvious fact that we do not possess a direct knowledge of this quality of the Universe. That is why an analysis of its features and limits always takes place by means of physical theories. These theories become the only way to perceive quantitative relations that occur in the physical world.

Therefore, we face the next boundary of our analyses. We cannot discuss ultimate boundaries of computability, but the limits of computability possibilities that result from a physical theory which is regarded as given. However, it may appear that such a far-reaching claim, namely the postulate of realizing infinity (energy, time) in finite sector of physical reality is not acceptable to every physical theory. To weaken slightly the last sentence, it is possible to restrict the considerations at least to commonly approved (not particularly exotic) physical theories. It occurs, though, that the above assumptions are not true. Two examples of physical theories allowing hypercomputability will be presented at this stage.

The first is the Newton mechanics. In 19th century P. Painlevé together with H. Poincaré proposed a particular analysis of an issue connected with the mechanics of heavenly bodies, that is the question of n-bodies. In the very issue of n-bodies a solution of an equation system of movements for n gravitational interacting bodies is sought. P. Painlevé and H. Poincaré opened a discussion not about the way to discover a particular solution, but about an analysis of qualities of these solutions. There is a crucial question whether there may exist such problem solutions that contain a singularity. The singularity as a solution has the quality such that its equation adopts infinite (not specified) values. It is obvious that a situation of this kind happens when two, from all the described by the problem, bodies collide. Yet the question arises whether the singularity may appear without any collision. The answer to this question was given by Z. Xia [19] in 1992. He claimed that for a problem of five bodies in the three-dimensional space there exist non-collision solutions. The Xia answer causes throwing one of the bodies to infinity in finite time. As it can be seen, the Newton mechanics allows finite realizations of infinity and potentially supports a possibility of calculations exceeding the limits of the Turing machine.

An obvious aim that appears at this moment is relating similar considerations to the physical theories that are regarded as currently valid. For this purpose we will use the theory of general relativity. There are such solutions of Einstein's equations in which there exists a time-like half-curve γ as well as a point p in spacetime such that the entire stretch γ is contained in the chronological past of p. Such spacetime structures (i.e. anti-de Sitter spacetimes) have been examined by physics with pointing out possible material systems that fulfill the required qualities (comp. [5]). Moreover, the descriptions of the usage of such systems to create computing systems [15] have been proposed. Summing up, the next of the analysed theories, the one which is regarded to be valid nowadays, allows conducting an infinite number of operations in limited time a precisely chosen observer.

The above results do not entitle us to accept the thesis that hypercomputability is possible in our world. They show, however, that the possibility of crossing the Turing machine barriers is, in the light of some physical theories, real.

As we can observe, a new cognitive situation is introduced. The boundaries of computability become valid only for a stated physical theory. Moreover they receive provisional and temporary character. When a physical theory regarded as the proper description of the Universe changes, there may occur a change in computability boundaries. The theory of computability gained additionally a relative character, this time in relation towards the physical theory assumed as a starting point in a construction of computing systems.

Conclusions

We shall try to determine the conclusions arising from the discussed results. From a mathematical point of view the structure of certain computability model is based on the qualification of a means of computability procedure construction. Though a description built on a finite dictionary results in a limited number of such procedures, other restrictions are not connected with it. The examples of models given above and exceeding Turing's limits prove that a formal description of models, considerably different from classical models in their computing power, is possible. Of course, we can built many of such models with their cardinality not exceeding $\aleph_0$. That is why an explanation of the choice of one has to arrive from a domain beyond mathematics.

Because the aim of the computability theory is a description of qualities of mechanically computable procedures, it is natural to turn to physics. It is physics that allows us to determine which of the proposed models are physically realized (which devices can be really constructed). Thus the criterion of differentiating problems into computable and noncomputable is moved to the domain of empiricism. The problems that solving can be described in the reality of the material world, may be regarded as computable in a completely intuitive way. Such a classification could become an absolute one. It means that the structure of the Universe separates explicitly the range of possibly computable problems from noncomputable issues, with no relativity possible.

However, the cognitive situation is different. Namely, we have no direct knowledge of the quality of the Universe in order to carry out the above-mentioned computability classification in an unquestionable way. Our perception of the material world is limited by an intermediary factor – a physical theory. Different aspects of reality can be contained in different theories, sometimes even in the same segment of the world several various theories are permitted, provided they are empirically consistent. An additional factor of theoretical physics variability is time that brings new research paradigms.

In this contest the cognition of the degree of computability of a certain issue must be perceived through a commonly approved physical theory. This new situation changes the way of thinking within the theory of computability. The problems absolutely and simply computable are out of the question. Now the relationship between a problem and some physical theory should be pointed out. A problem P can be computable in relation towards quantum mechanics and noncomputable to Newton's mechanics.

Let us notice that such a view on the theory of computability refers not only to differentiation of problems into computable and noncomputable. The relativity towards a physical theory influences also the issues of complexity. It is worth pointing out that an adoption of a factor of time to be a continuous parameter allows in appropriate models to solve any number of operations (in suitably short time) in constant time. So the practical complexity (realization time, not a number of steps) of problems that are regarded to be intractable due to their classically understood complexity, may be reduced in this approach.

What is then the role of mathematics and logic in the theory of computability? Traditionally the sciences will act as the language that expresses research issues in this field. Moreover, protecting cohesion and completeness of introduced models, they will guarantee their correctness. But a choice of model and establishing limits of computing power will stay beyond the domain of mathematics, which is merely a research tool now.

The end of this paper will be presented in the form of concise thesis which results from all the above considerations.

The issue of computability is more a problem of ontology than epistemology. *It is a shape of physical reality, not any constructions a priori of human reason, that decides which problems are computable.*

Researches of the limits of computability are relative towards intermediary physical theories. Therefore, having no direct insight into the Universe ontology, *we have to base our researches of the computability theory on comprehensive physical theories.*

Considering computations beyond the limits of the Turing model is justified as a reasonable research program. The given examples of physical theories allowing hypercomputability (Newton's mechanics, the general relativity theory) show that exceeding the limits designed by a Turing machine model is not – at least potentially – unlikely.[2]

References

[1] H. P. Barendrengt. *The Lambda Calculus*, North-Holland, 1981.

[2] B. J. Copeland. The Church-Turing Thesis. In J. Perry and E. Zalta, (eds), *Stanford Encyclopedia of Philosophy*, 1996.

[2] The question of noncomputability in nature is a very controversional one. We have results of Pour-El and Richards [12] which suggest an existance of some physical phenomena beyond Turing computability. Contrary, others (e.g. [11]) reject a possibilty of noncomputable devices in nature pointing out an artificial and nonsmooth character of the mentioned examples.

[3] B. J. Copeland. Even Turing machines can compute uncomputable functions. In C. S. Calude, J. Casti, and M. J. Dinneen, (eds), *Unconventional Models of Computation*, Springer-Verlag, 1998.

[4] R. O. Gandy. Church's Thesis and principles of mechanisms. In J. Barwise, J. Keisler and K. Kunen, (eds), *The Kleene Symposium*, North--Holland, 1980.

[5] G. Etesi and I. Nemeti. Non-Turing computations via Malament-Hogarth space-times. *Int. J. Theor. Phys.* 41:341-370, 2002.

[6] A. A. Markow. Theory of algorithms (in Russian). *Trudy Matiematiczeskogo Instituta Stieklova*, 38: 176:189, 1951.

[7] C. Moore. Recursion theory on the reals and continuous-time computation. *Theoretical Computer Science*, 162:23-44, 1996.

[8] R. Murawski. *Funkcje rekurencyjne i elementy metamatematyki*, Wydawnictwo Naukowe UAM, 1991.

[9] J. Mycka, J. F. Costa. Real recursive functions and their hierarchy. Submitted to *Journal of Complexity*, 2003.

[10] P. Odifreddi. *Classical Recursion Theory*, North-Holland, 1989.

[11] R. Penrose. *The Emperor's New Mind*, Oxford University Press, 1989.

[12] M. B. Pour-El, J. I. Richards. Noncomputability in analysis and physics. *Adv. Math.* 48:44-74, 1983.

[13] L. A. Rubel. Some mathematical limitations of the general-purpose analog computer. *Advances in Applied Mathematics*, 9:22-34, 1988.

[14] L. A. Rubel. The extended analog computer. *Advances in Applied Mathematics*, 14:39-50, 1993.

[15] O. Shagrir, I. Pitowsky. Physical Hypercomputation and the Church--Turing Thesis. *Minds and Machines*, 13: 87-101, 2003.

[16] C. Shannon. Mathematical theory of the differential analyzer. *J. Math. Phys. MIT*, 20:337-354, 1941.

[17] R. Soare. The history and concept of computability. In E. Griffor, (ed), *Handbook of Computability Theory*, Elsevier, 1999.

[18] A. Turing. On computable numbers, with application to the Entscheidungsproblem. *Proc. London Math. Soc.*, Ser. 2, 42: 230-265, 1936.

[19] Z. Xia. The existence of noncollision singularities in Newtonian systems. *The Annals of Mathematics*, 135(3):411-468, 1992.

Jerzy Mycka
Institute of Mathematics
University of Maria Curie-Sklodowska
Lublin, Poland

Dariusz Surowik
University of Białystok

TEMPORAL LOGIC APPROACH TO EXTENSIVE GAMES[1]

A mathematical n-players game model can be represented in a normal or extensive form. A normal form representation of games is ideal to represent situations where players make one choice and move simultaneously. An extensive form provides an explicit description of a strategic interaction by specifying a physical order of play, actions available to players each time they get to choose, and eventual payoffs for each player for any sequence of choices. For these reasons the extensive form provides a richer environment to study interesting questions such as rivalry, repeated interaction, etc.

Obviously every normal form of the game can be represented in an extensive form, but it is more natural and simpler just to write the normal form. It is less obvious that every extensive form of the game can be written in a normal form. This translation is possible if we note that a strategy in an extensive form game is not just a move or sequence of moves, but rather is a complete contingency plan. A strategy for a player i must specify what the player will do at every node or what information set the player has.

Games in extensive form

In this section we will study extensive form games. The basic notion of the extensive form games theory is a *game tree* notion. It contains nodes and branches. Nodes represent decision points where only one player has to make a decision. Branches represent possible choices available for player. A game tree includes all alternative actions that can be taken by all players

[1] The research reported in this paper is a part of the project entitled *Undecidability and Algorithmic Intractability in the Social Sciences* supported by the Polish Ministry of Science, grant no. 2 H01A 030 25.

ISBN 83-7431-015-4 ISSN 0860-150X

and all possible outcomes. If we construct a game tree, we have to obey the following rules:

- at least one branch leads from each decision node,
- only one branch leads to a decision node.

Example 1

Let us consider the following examples of structures:

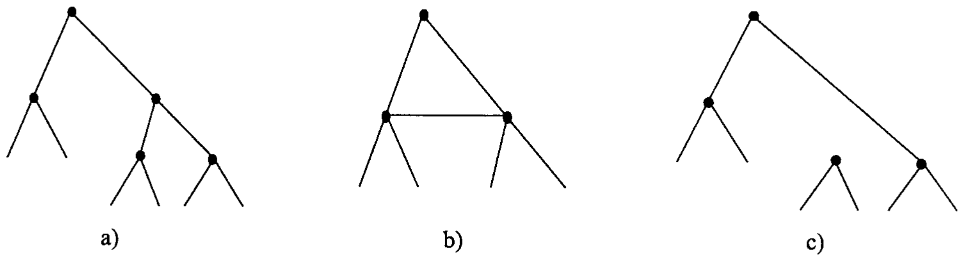


Fig. 1.

In case a) we have a game tree. In case b) tree does not form a game tree, because in the structure of the tree we have a closed path. In case c) we do not have a game tree, because there is not a path between two nodes.

Definition 1

A rooted tree is a pair $\langle T, \longmapsto \rangle$, where T is a set of nodes and $\longmapsto$ is a binary relation on T satisfying the following conditions:

- there is a distinguished node t_0 ($\in T$) (it is called the *root*), such that no immediate predecessors,
- for every node $t \in T \setminus \{t_0\}$ there exists a unique path from t_0 to t.

A *terminal node* is a node which has no immediate successors. Let $L(T)$ denote the set of terminal nodes.

Example 2

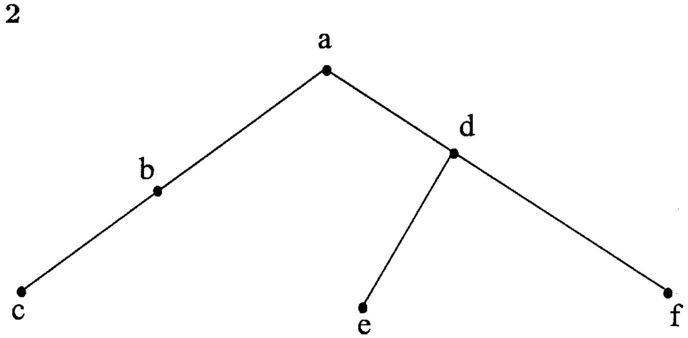


Fig. 2. A rooted tree with the root a and terminal nodes c, e, f

Let us remark that in the normal form game all parts of the game other than the strategies are removed (for this reason it is also called the strategic form). For a given game in a normal (strategic) form

		PLAYER II	
		C	D
PLAYER I	A	1,1	5,0
	B	2,3	2,3

Fig. 3.

we can consider a few extensive forms. Two extensive forms of the normal form showed in Fig. 3 could be formed in the following way:

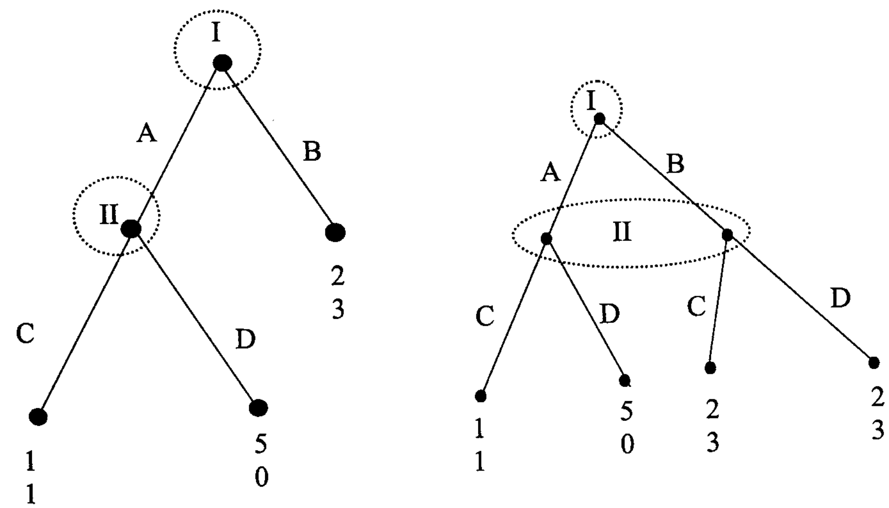


Fig. 4. Two extensive form versions of the same normal form game

If a game is played sequentially, the extensive form allows us to see exactly how the game could be played out.

The knowledge which a player has after every move is essential from a player standpoint. Sometimes in games moves are taken in a sequence and every player observes every event that takes place until that player has to take an action. However, there are games whose rules are such that the players do not obtain the so-called perfect information on the decisions of the other players. This kind of situation usually takes place in the card games when the first move is random.

Now we will explain the differences between perfect information games and imperfect information games. Let us consider the following example:

Example 3
Let us take a game with the following game tree:

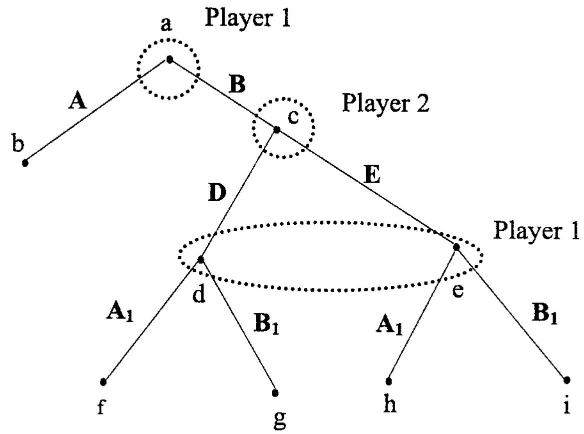


Fig. 5.

A circle (dotted line) denotes information sets. It surrounds (or connects) the nodes that are *indistinguishable* from the decision maker's standpoint. In the above case in the node d or the node e the player 1 does not know the player 2's decision, and therefore he has imperfect information.

Definition 2
An *information set* is a collection of the decision nodes so that:

1) The same player is mapped to all these nodes,
2) If the play of the game reaches a node in this collection, the player does not know which node has been reached.

A *game of perfect information* is a game in which there is no information set with multiple nodes[2]. If there are multiple nodes information sets in a game, then we have a game of imperfect information.

Now we give some conditions for a finite extensive form with perfect information.

[2] At any node each player knows the entire history of play.

Definition 3
A ***finite extensive form with perfect information*** is a tuple: $\langle T, \longmapsto, N, \xi \rangle$, where:

- $\langle T, \longmapsto \rangle$ is a finite rooted tree,
- $N = \{1, \ldots, n\}$ is a set of players,
- $\xi : (T \setminus L(T)) \to N$ is a function that associates with every non-terminal or decision node of the player who moves at that node.

Given an extensive form, we obtain a perfect information game by adding for every player $i \in N$, for every terminal node $t \in L(T)$ a payoff or utility function $[u(t) = (u_1(t), \ldots, u_n(t))]$.

Temporal logic of branching time

The idea of temporal logic of branching time was given by A. N. Prior[3]. One of the main motivations of the construction of temporal logic of branching time was a wish for creation of indeterministic temporal logic. Arguments on determinism were rejected by modification of the structure of time. The basic system of the temporal logic of branching time is Nino Cocchiarella's system called „*CR*"[4]. In the *CR* system a relation of temporal succession is transitive. Because no other conditions are imposed upon the earlier-later relation, then the symmetry of the past and the future[5] is possible in *CR*. In the other systems of temporal logic of branching time there are additional conditions imposed on the earlier-later relation. A left--linearity property of the earlier-later relation is necessary in the K_b[6] system, for example.

An example of the structure of time linear in the past is presented below (Fig. 6).

As we see, the past has no alternatives and it is determined, but the future is open and there are a lot of ways of its realization. Among all possible futures only one is realized. It is called *actual future*.

[3] A. N. Prior, *Past, Present and Future*, Oxford University Press, 1967.
[4] A. N. Prior, *Past, Present and Future*, Oxford University Press, 1967, Appendix A.
[5] The system *CR* is often considered as a model of ideas in the contemporary physics on real time R. P. McArthur, *Tense logic*, Dordrecht 1976, p. 39.
[6] N. Rescher, A. Urquhart, *Temporal Logic*, Wien, New York, 1971, chapter 4.

Example 4

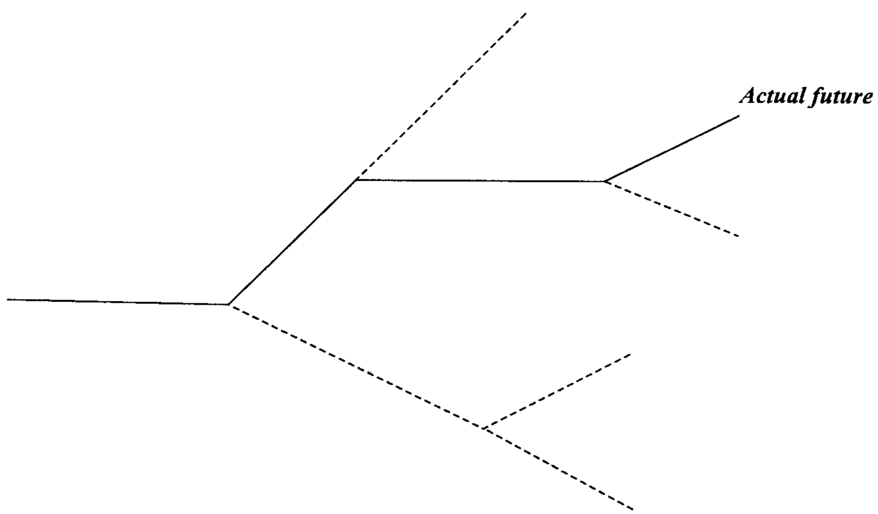


Fig. 6.

Interpretation of extensive games in terms of temporal logic of branching time

The logical foundations of game theory have been a subject of consideration of many scientific papers. Usually these considerations are formulated in terms of modal or epistemic logic. However games are specific processes in which dynamic interactions between the players occur. If we consider a notion of process, we have to consider its temporal context. It seems that a natural way of interpretation of games is its interpretation in terms of temporal logic. Extensive games can be modeled in terms of temporal logic of branching time if we add a notion of agent to the semantics and define a notion of prediction.

Definition 4

A branching-time frame with agents[7] (*BTA-frame* for short) is a tuple $\langle T, \prec, N, \{R_i\}_{i \in N} \rangle$, where:

- T – a set of nodes,

[7] Bonanno G., *Branching time, perfect information games and backward induction*, 1999.

- $\prec$ – a binary relation on T (the precedence relation) satisfying the following conditions:
  A1) if $t_1 \prec t_2$, then $t_2 \not\prec t_1$ (antisymmetry)
  A2) if $t_1 \prec t_2$ and $t_2 \prec t_3$, then $t_1 \prec t_3$ (transitivity)
  A3) if $t_1 \prec t_3$ and $t_2 \prec t_3$ then $t_1 = t_2$ or $t_1 \prec t_2$ or $t_2 \prec t_1$ (left linearity),
- $N = \{1, \dots, n\}$ is a finite set of agents,
- for any $i \in N$, R_i is a binary relation on T, such that: if $t_1 R_i t_2$, then $t_1 \prec t_2$ (R_i is subrelation of $\prec$).

Properties A1-A3 constitute the definition of branching time. In particular, the left linearity property limits a class of frames to the frames, where at any node the future has alternatives and the past is unique.

The interpretation of $t_1 R_i t_2$ is as follows: at node t_1 agent i can make a decision, which leads from t_1 to t_2. It is possible, that for some agent i and for some node t, the set $R_i(t) \stackrel{\text{def}}{=} \{t' \in T : tR_it'\}$ is empty. In this case agent i does not have any actions available at node t.

Definition 5

For a given a BTA-frame *prediction* is a binary relation $\prec_P$ on T, satisfying the following conditions:

P1) if $t_1 \prec_P t_2$, then $t_1 \prec t_2$ ($\prec_P$ is subrelation of $\prec$),
P2) if $t_1 \prec_P t_2$ and $t_2 \prec_P t_3$, then $t_1 \prec_P t_3$ (transitivity),
P3) if $t \prec t_1$ for some t_1, then $t \prec_P t_2$ for some t_2,
P4) if $t_1 \prec t_2$, $t_2 \prec t_3$ and $t_1 \prec_P t_3$, then $t_1 \prec_P t_2$ and $t_2 \prec_P t_3$.

The condition P1 shows that the relation of prediction is subrelation of $\prec$. The conceivable future is a subset of the set of all future nodes. Let us remark, that we do not assume that the conceivable future is unique for a given node. We do not require $t' = t''$, $t' \prec_P t''$, $t'' \prec_P t'$ in case $t \prec_P t'$ and $t \prec_P t''$. Moreover, we do not assume that the predictable future for a given node is a proper subset of the conceivable future. P2 (transitivity) is a natural condition for a notion of prediction[8].

Every $t \in T$ should be thought of as a complete description of the world. Sets of nodes represent propositions. We introduce a formal notation and a notion a model for a correct interpretation.

Let us consider a language of propositional logic with the following specific operators: $G, H, G_p, H_p, \rhd_i$.

[8] More detailed discussion on these conditions we find in G. Bonanno, *Branching time, perfect information games and backward induction*, 1999.

The interpretation of the operators is as follows:

$G\alpha$ – it is going to be the case in every conceivable future that α,
$H\alpha$ – it has always been the case that α,
$G_p\alpha$ – it is going to be the case in every predictable future that α,
$H_p\alpha$ – it has always been the case at every past node at which the current node was predicted that α,
$\rhd_i\alpha$ – no matter what action agent i takes, it will be the case that α.

The operators G and H are the standard operators of the temporal logic of branching time. The operators G_p and H_p are specific operators introduced by G. Bonanno[9]. The operators $\rhd_i$ (for any $i \in N$) are action's operators.

Formal language

Alphabet:

- a countable set of propositional letters S,
- connectives: $\neg, \Rightarrow$,
- temporal operators: G, H, G_p, H_p,
- action operators $\rhd_i$,
- parentheses:), (.

The set of the sentences is defined as follows:

Definition 6

The set of the sentences is the smallest set $\mathbb{Z}$, such that:

- $S \subseteq \mathbb{Z}$,
- if $\alpha, \beta \in \mathbb{Z}$ then $\neg\alpha, (\alpha \Rightarrow \beta), G\alpha, H\alpha, G_p\alpha, H_p\alpha, \rhd_i\alpha \in \mathbb{Z}$.

We adopt the following definitions:

Definition 7

$$\begin{aligned}(\alpha \vee \beta) &\equiv (\neg\alpha \Rightarrow \beta),\\ (\alpha \wedge \beta) &\equiv \neg(\alpha \Rightarrow \neg\beta),\\ (\alpha \Leftrightarrow \beta) &\equiv \neg[(\alpha \Rightarrow \beta) \Rightarrow \neg(\beta \Rightarrow \alpha)],\\ F\alpha &\equiv \neg G \neg\alpha,\\ P\alpha &\equiv \neg H \neg\alpha,\\ F_p\alpha &\equiv \neg G_p \neg\alpha,\\ P_p\alpha &\equiv \neg H_p \neg\alpha.\end{aligned}$$

[9] Bonanno G., *Branching time, perfect information games and backward induction*, 1999.

By adding a function $V : S \to 2^T$ to a given BTA-frame we obtain a model $\mathfrak{M}$ based on this frame. Validation for formulas is as follows:

Definition 8

a) $\mathfrak{M}, t \models \alpha \equiv t \in V(\alpha)$, if $\alpha \in X$,
b) $\mathfrak{M}, t \models \neg\alpha \equiv$ not $\mathfrak{M}, t \models \alpha$,
c) $\mathfrak{M}, t \models (\alpha \Rightarrow \beta) \equiv$ if $\mathfrak{M}, t \models \alpha$, then $\mathfrak{M}, t \models \beta$,
d) $\mathfrak{M}, t \models G\alpha \equiv$ for every t' such that $t \prec t'$ holds $\mathfrak{M}, t' \models \alpha$,
e) $\mathfrak{M}, t \models H\alpha \equiv$ for every t' such that $t' \prec t$ holds $\mathfrak{M}, t' \models \alpha$,
f) $\mathfrak{M}, t \models G_p\alpha \equiv$ for every t' such that $t \prec_p t'$ holds $\mathfrak{M}, t' \models \alpha$,
g) $\mathfrak{M}, t \models H_p\alpha \equiv$ for every t' such that $t' \prec_p t$ holds $\mathfrak{M}, t' \models \alpha$,
h) $\mathfrak{M}, t \models \rhd_i\alpha \equiv$ for every t' such that tR_it' holds $\mathfrak{M}, t' \models \alpha$.

Theorem 1

A finite extensive form with perfect information is a special case of a BTA frame[10].

Game model

The theorem 1 shows that we can consider a finite extensive form with perfect information as a special case of a BTA frame. To view a perfect information game as a model we need so, that the set of sentences include sentences of the form $(u_i = q)$, where $i \in N$, and $q \in Q$. The interpretation of the sentences of the form $(u_i = q)$ is: *player i's payoff is*. The sentence of the form $(q_1 \leq q_2)$ we interpret as: *the payoff q_1 is less than or equal to the payoff q_2*.

Definition 9

Let $\mathfrak{J}$ be the BTA frame corresponding to a given perfect information game. A game model $\mathfrak{M}$ is a model based on $\mathfrak{J}$ obtained by adding to $\mathfrak{J}$ a valuation $V : s \to 2^T$ satisfying the following conditions:

- if p $(\in S)$ is the sentence of the form $(q_1 \leq q_2)$, then $V(p) = T$, if $(q_1 \leq q_2)$, and $V(p) = \emptyset$ in the otherwise,
- if p $(\in S)$ is the sentence of the form $(u_i = q_2)$, then $V(p) = \{t \in L(T) : u_i(t) = q\}$.

[10] G. Bonanno, *Branching time, perfect information games and backward induction*, 1999.

We have the following conclusions: if $\mathfrak{M}$ is a game model, then $\forall_{t \in T}\ \mathfrak{M}, t \models (q_1 \leq q_2)$ if q_1 is less than or equal to the q_2 and $\mathfrak{M}, t \models \neg(q_1 \leq q_2)$ otherwise. In the model $\mathfrak{M}$, at node t holds $\mathfrak{M}, t \models (u_i = q)$, if t is a terminal node, such that $u_i(t) = q$. In the model $\mathfrak{M}$, at node t holds $\mathfrak{M}, t \models \neg(u_i = q)$ in a case, if t is either a decision node or terminal node such that $u_i(t) \neq q$.

In our formal language we can consider the truth of specific formulas in a model. These formulas describe some specific properties of games. Le us consider the following formula, for example:

I1) $F_p(u_i = q) \Rightarrow \rhd_i [((u_i = r) \vee F_p(u_i = r)) \Rightarrow (r \leq q)]$

The interpretation of this formula is as follows: if it is predictable that player i's payoff will be q then, no matter what decision the player i makes, if his payoff is r, or it is predictable that it will be r, then r is not greater than q.

Let us consider another formula:

I2) $$[F_p(u_i = q) \wedge (F_p(u_i = s) \Rightarrow (q \leq s))] \Rightarrow$$
$$\Rightarrow \rhd_i \{[(u_i = r) \vee (F_p(u_i = r) \wedge (F_p(u_i = s) \Rightarrow (r \leq s)))] \Rightarrow (r \leq q)\}$$

The interpretation of it is: if, according to the prediction, player i's payoff will be at least q, then, no matter what decision player i makes, if his payoff is r, or is predicted to be at least r, then r is not greater than q.

Both formulas I1 and I2 characterize the backward induction algorithm[11] for generic games in terms of temporal logic of branching time[12].

The backward induction algorithm can be used to solve finite extensive games with perfect information. What about infinite extensive games? Let us consider temporal logic based on intuitionistic propositional logic.

[11] *Backward induction algorithm* is as follows: Suppose the initial decision node is K steps removed from the terminal nodes i.e. the maximum number decision nodes between the initial node and any terminal nodes is K.
Step 1: At any final decision node, every decision–making player chooses a move that maximizes her payoff.
Step 2: At a penultimate decision node, every decision-making player anticipates step 1 and chooses a payoff maximizing move.
...
Step k: At decision nodes k steps removed from a terminal node, every decision-making player anticipates step $k = 1$, for each $k = 3, \ldots, K$ and chooses a payoff maximizing move.

[12] Proof of this fact we find in the G. Bonanno, *Branching time, perfect information games and backward induction*, 1999.

Intuitionistic temporal logic

Alphabet:
- set of propositional letters: Ψ,
- intuitionistic unary connective: $\neg$,
- intuitionistic binary connectives: $\wedge$, $\vee$, $\Rightarrow$, $\Longleftrightarrow$,
- temporal operators: G, H, F, P,
- parentheses:), (.

Notation

I – a non-empty set of indexes of state of knowledge,

$T_i \ (i \in I)$ – a non-empty set of moments of time in a state of knowledge indexed by i,

$R_i \ (\subseteq T_i \times T_i)$ – a binary relation on T_i,

$\mathcal{T}_i \ (= \langle T_i, R_i \rangle)$ – a time in a state of knowledge indexed by i,

$\mathcal{T} = \bigcup_{i \in I} \mathcal{T}_i$ – a set of all moments of time,

$R \ (= \bigcup_{i \in I} R_i)$ – a binary relation on the set of all moments of time,

$V_i \ (\subseteq T_i \times 2^{\Psi})$ – a function mapping to elements $t \ (\in T_i)$ subsets of the set of propositional letters,

$\wp = \{V_i : \in I\}$ – a class of function V_i,

$m_i \ (= \langle T_i, R_i, V_i \rangle)$– a state of knowledge indexed by i,

$\mathfrak{M}_{(\mathcal{T},\wp)} = \{\langle T_i, R_i, V_i \rangle : V_i \in \wp,\ i \in I\}$, then $\mathfrak{M}_{(\mathcal{T},\wp)} = \{m_i : i \in I\}$. $\mathfrak{M}_{(\mathcal{T},\wp)}$ is a model based on time $\mathcal{T}$ and class of functions $\wp$.

Between elements of a model $\mathfrak{M}_{(\mathcal{T},\wp)}$ we introduce a relation $\leq$ $(\subseteq \mathfrak{M}_{(\mathcal{T},\wp)} \times \mathfrak{M}_{(\mathcal{T},\wp)})$.

Definition 10

For any $i, j \in I$:

$$m_i \leq m_j \quad = \quad (T_i \subseteq T_j \text{ and } R_i \subseteq R_j \text{ and } \forall_{t \in T_i} V_i(t) \subseteq V_j(t)).$$

„$m_i \leq m_j$" means that the state of knowledge m_j is not smaller than the state of knowledge m_j.

Let us consider the example which shows various ways to obtain new states of knowledge. A state of knowledge may not be enlarged to a less one in several cases. One of them is when in the new state of knowledge we are able to describe events which were not known in the smaller one. Another case is when we obtain a new knowledge about the structure of time. One more case is when in a new state of knowledge the relation between time points is changed.

Example 5

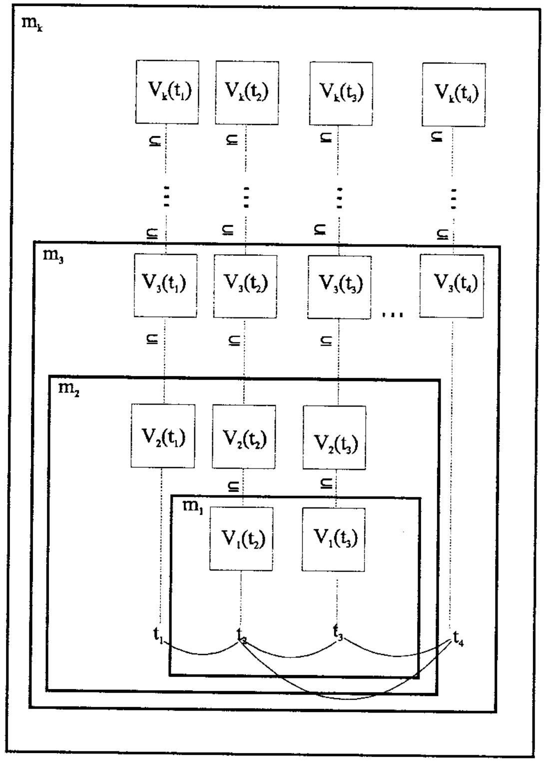


Fig. 7.

Remark

m_i^* (where $i \in I$) means any m_j ($\in \mathfrak{M}_{(\mathcal{T},\wp)}$) such that $m_i \leq m_j$.

Definition 11[13]

For a model $\mathfrak{M}_{(\mathcal{T},\wp)}$, state of knowledge m_i $(= \langle T_i, R_i, V_i \rangle)$, element t $(\in T_i)$, a formula α $\mathfrak{M}_{(\mathcal{T},\wp)} \models \alpha[t, m_i]$, is defined by the following conditions:

[13] An axiomatization of the minimal intuitionistic temporal logic was given by Surowik D. in the *Tense Logic Without The Principle of The Excluded Middle*, Topics in Logic, Informatics and Philosophy of Science, Białystok, 1999.

a) $\mathfrak{M}_{(\mathcal{T},\wp)} \models \alpha[t, m_i] \quad \equiv \alpha \in V_i(t)$, if $\alpha \in \Psi$,

b) $\mathfrak{M}_{(\mathcal{T},\wp)} \models \neg\alpha[t, m_i] \quad \equiv \forall_{m_i^* \in \mathfrak{M}_{(\mathcal{T},\wp)}} \mathfrak{M}_{(\mathcal{T},\wp)} \not\models \alpha[t, m_i^*]$,

c) $\mathfrak{M}_{(\mathcal{T},\wp)} \models (\alpha \vee \beta)[t, m_i] \quad \equiv \mathfrak{M}_{(\mathcal{T},\wp)} \models \alpha[t, m_i]$ or $\mathfrak{M}_{(\mathcal{T},\wp)} \models \beta[t, m_i]$,

d) $\mathfrak{M}_{(\mathcal{T},\wp)} \models (\alpha \wedge \beta)[t, m_i] \quad \equiv \mathfrak{M}_{(\mathcal{T},\wp)} \models \alpha[t, m_i]$ and $\mathfrak{M}_{(\mathcal{T},\wp)} \models \beta[t, m_i]$,

e) $\mathfrak{M}_{(\mathcal{T},\wp)} \models (\alpha \Rightarrow \beta)[t, m_i] \quad \equiv \forall_{m_i^* \in \mathfrak{M}_{(\mathcal{T},\wp)}} (\mathfrak{M}_{(\mathcal{T},\wp)} \not\models \alpha[t, m_i^*]$ or $\mathfrak{M}_{(\mathcal{T},\wp)} \models \beta[t, m_i^*])$,

f) $\mathfrak{M}_{(\mathcal{T},\wp)} \models F\alpha[t, m_i] \quad \equiv \exists_{t_1 \in T_i}$ (such that tR_it_1 and $\mathfrak{M}_{(\mathcal{T},\wp)} \models \alpha[t_1, m_i]$),

g) $\mathfrak{M}_{(\mathcal{T},\wp)} \models G\alpha[t, m_i] \quad \equiv \forall_{m_i^* \in \mathfrak{M}_{(\mathcal{T},\wp)}} \forall_{t_1 \in T_i^*}$ (if $tR_i^*t_1$, then $\mathfrak{M}_{(\mathcal{T},\wp)} \models \alpha[t_1, m_i^*]$),

h) $\mathfrak{M}_{(\mathcal{T},\wp)} \models P\alpha[t, m_i] \quad \equiv \exists_{t_1 \in T_i}$ (such that t_1R_it and $\mathfrak{M}_{(\mathcal{T},\wp)} \models \alpha[t_1, m_i]$),

i) $\mathfrak{M}_{(\mathcal{T},\wp)} \models H\alpha[t, m_i] \quad \equiv \forall_{m_i^* \in \mathfrak{M}_{(\mathcal{T},\wp)}} \forall_{t_1 \in T_i^*}$ (if $t_1R_i^*t$, then $\mathfrak{M}_{(\mathcal{T},\wp)} \models \alpha[t_1, m_i^*]$).

In the language of the intuitionistic temporal logic the argument on determinism based on the principle of excluded middle is formulated as follows:

$$F\alpha \vee F\neg\alpha$$

The formula $F\alpha \vee F\neg\alpha$ is not a tautology of intuitionistic temporal logic. A countermodel for formula $Fp \vee F\neg p$ is showed on Fig. 8.

As we can see, the formula Fp is not true in our model in the state of knowledge m_1 at moment t, because in the state m_1 there is not a moment of time t' later than t, so that the sentence p is true at t'. The formula $F\neg p$ is not true in the state of knowledge m_1 at moment t either. The necessary condition for the truth of the sentence $F\neg p$ is: the sentence p is false at every moment of time later than t in every state of knowledge not smaller than the state of knowledge m_1. As we can see, our model does not satisfy this condition.

Infinite branching time structures are proper for semantic considerations in the temporal logic of branching time based on intuitionistic logic. We can consider an intuitionistic temporal logic of branching time if we add the following formulas to the axioms of the minimal intuitionistic temporal logic:

B1) $FF\alpha \Rightarrow F\alpha$,

B2) $G\alpha \Rightarrow GG\alpha$,

B3) $(P\alpha \wedge P\beta) \Rightarrow [P(\alpha \wedge \beta) \vee P(P\alpha \wedge \beta) \vee P(\alpha \wedge P\beta)]$.

Example 6[14]

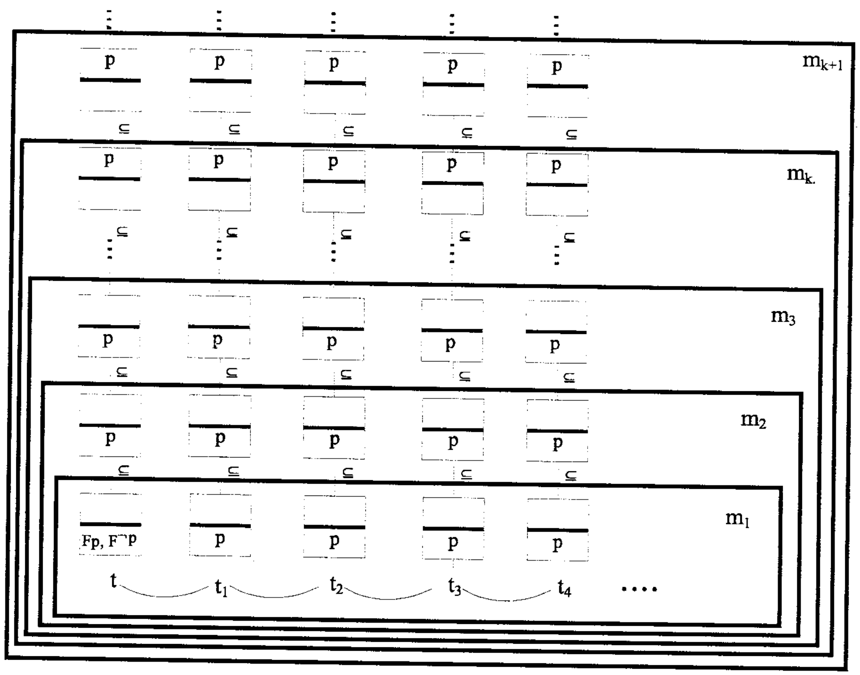


Fig. 8.

These formulas characterize the class of time branching in the future in terms of intuitionistic temporal logic[15].

We can imagine that in the terms of intuitionistic temporal logic we analyze a structure of time very similar to the structure of the infinite extensive game. For example:

[14] If a formula is above the horizontal line in the square mapped for a moment t in a given state of knowledge, then we interpret that the formula is true at a moment t in this state of knowledge. Otherwise, we mean that the formula is false at this moment in this state.

[15] The various properties of the earlier-later relation in the intuitionistic temporal logic are discussed in Surowik D., *Some Remarks about Intuitionistic Tense Logic*, On Leibniz's Philosophical Legacy, Białystok 1997.

Example 7

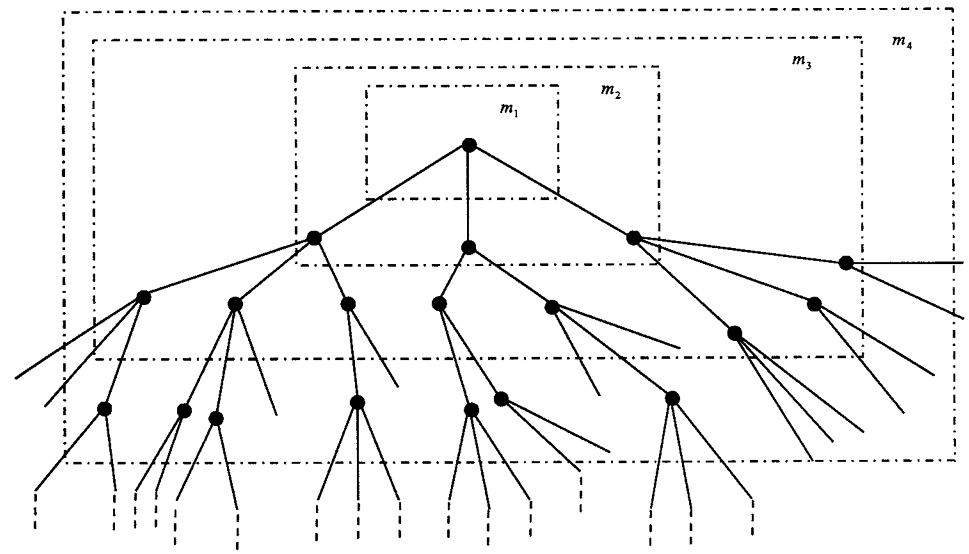


Fig. 9.

The open problem is:

Is a modification of intuitionistic temporal logic of branching time (by adding agents and payoffs to the semantics) possible for the analyses of infinite indeterministic extensive games?

References

1. J. van Benthem, *Logic and game theory*, Proceedings of LLC9, 2000
2. G. Bonanno, *Branching time, perfect information games and backward induction*, Third Conference on Logic and the Foundations of Game and Decision Theory, Turin, 1999
3. G. Bonanno, *Modal logic and game theory, two alternative approaches*, Strategic Rationality in Economics, 1999
4. G. Bonanno, *Prediction in branching time logic*, 2000
5. R. P. McArthur, *Tense logic*, Dordrecht 1976
6. N. Prior, *Past, Present and Future*, Oxford University Press, 1967
7. N. Rescher, A. Urquhart, *Temporal Logic*, Wien, New York, 1971
8. D. Surowik, *Some Remarks about Intuitionistic Tense Logic*, On Leibniz's Philosophical Legacy, Białystok, 1997

9. D. Surowik, *Tense Logic Without The Principle of The Excluded Middle*, Topics in Logic, Informatics and Philosophy of Science, Białystok, 1999

Dariusz Surowik
Department of Logic, Informatics and Philosophy of Science
University of Białystok
e-mail: surowik@hum.uwb.edu.pl

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 7 (20) 2004

Joanna Domurat
Warsaw School of Economics

Tadeusz Tyszka
Leon Kozminski Academy
of Entrepreneurship and Management

HISTORIC AND CONTEMPORARY CONTROVERSIES ON THE CONCEPT OF UTILITY[1]

Abstract. The concept of utility was first introduced in the demand theory, together with the assumption of decreasing marginal utility. Later, postulates of consumer rational preferences and their violations were debated. The classic concept of utility, which was based exclusively on consumer preferences has recently been questioned by Kahneman and his colleagues, who introduced the concept of experienced utility. In the context of decisions under risk, for a long time the model of maximizing the expected utility has been accepted. However, some hardship to explain paradoxes were also debated. An alternetive model, known as the prospect theory, includes basic changes concerning the valuation of outcomes and probabilities by the decision maker. This "psychological" theory has turned out to have several appealing implications for various areas of economics.

1. Demand theory, the assumption of decreasing marginal utility, and postulates of consumer rational preferences

Utility is a term which, for a long, time has been used in connection with the desire to possess goods or services. For economists utility has been a subjective measure of value which differs from the value of goods expressed as a price on the market. For example, Adam Smith in his book "The Wealth of Nations" (1776/1986) states that the value of goods expressed by their price is distinct from their value in the sense of utility. Water, for example, has high utility but a low price. Diamonds, on the contrary, have low utility but a high price. Ricardo put it similarly in his book "Principles of Political

[1] Granted by KBN Nr 2 H01F 022 24.

ISBN 83-7431-015-4 ISSN 0860-150X

Economy and Taxation" (1817/1957). For him utility is not measured by the price on the market. However, goods which are exchangeable must be to some extent useful. Useless things are not exchanged.

Classical economists could not explain the full relationship between utility and demand on the market because explanation required analysis of marginal values. Many economists in the mid-19th century begun such analyses (eg. H. H. Gossen, W. S. Jevons, C. Menger, A. Marshall). (Stankiewicz 2000). For them consumer utility was a function which assigns values $u(y)$ for commodity bundles y. The term marginal utility is then an increment of utility as the result of consuming the successive unit of a commodity.

Economists noticed that consumption of each successive unit of a commodity brings us less utility. This is the law of decreasing marginal utility which has been till now one of the fundamentals of economics.

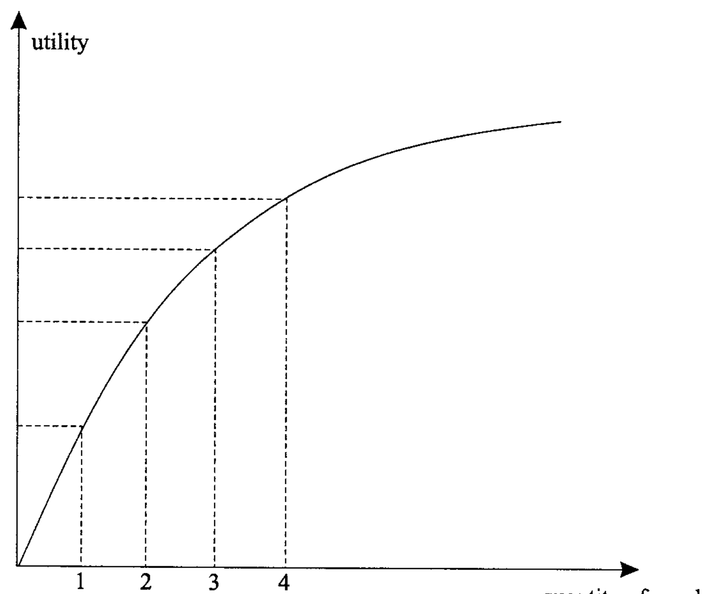


Fig. 1. Consumer utility function for one commodity

Utility then became a measure of consumer satisfaction, a measurable psychological magnitude. Total consumer utility was the sum of the utilities of every good consumed. This approach involved two basic problems: how to measure the utility and whether interpersonal comparisons of utility are possible.

Many economists were aware of these problems. For example Jevons (1871/1970) states that it is difficult to conceive a unit of utility. On the one hand, he regarded every man as distinct and noticed that it was not

possible to find a common denominator of feelings for two individuals. On the other hand, utility in his analysis is treated as a measurable value, and is compared and aggregated. This shows that in spite of many attempts, economists of that time failed to solve the problems of defining utility. The interesting thing is that these problems resulted in difficulties in developing a complete theory of the consumer. The producer theory for economists was easier because profit and other economic values connected with the production were measurable.

The problems connected with utility were not solved until the 1930s. In 1934 Hicks and Allen described how demand theory and utility can be based on consumer preferences alone. This work was later developed and began to dominate in microeconomic analysis. This meant that it was no longer necessary to create any psychological measure of utility. The theory goes as follows.

There is a set X comprising elements x_i, which can be quantities of any goods or bundles of these goods, e.g. a bar of chocolate, a loaf of bread or a ticket to a cinema. We choose from this set two elements and ask consumer which one he prefers. Notice that at this stage the consumer neglects the prices of the goods. By indicating the good he prefers, he sets a preference relation between these two goods. The week preference relation $x_1 \succcurlyeq x_2$ means that x_1 is at least as good as x_2. We can also define the strict preference relation $x_1 \succ x_2$ which means that the consumer prefers x_1 to x_2. The indifference relation $x_1 \sim x_2$ means that the consumer is indifferent as to which commodity to choose.

We assume that customer preferences are rational. This means that they possess the following two properties:

1) Completeness. This assumption means that a preference relation is set for any two alternatives from the set X, that is:

   for all $x_i, x_j \subset X$ at least one of relation holds: 1) $x_i \succcurlyeq x_j$ 2) $x_i \preccurlyeq x_j$

2) Transitivity. This assumption means that there is a stable hierarchy in the choices of elements from the set X.

   for all x_i, x_j, x_k: if $x_i \succcurlyeq x_j$ and $x_j \succcurlyeq x_k$ then: $x_i \succcurlyeq x_k$.

The above two postulates are not straightforward if we relate them to our everyday decisions. It is not always simple to set the preferences between two alternatives. Completeness assumes that we can make such a decision concerning any two alternatives. The other postulate is even stronger. It says that the preferences revealed in the pairwise choices will never circle and are ordered in a stable and consistent hierarchy.

The above postulates are very important in economics as they allow us to define the utility function[2]. Utility is based on rational preferences revealed by the consumer. Utility function assigns numerical values to elements of X. The commodity which is preferred, has higher value of utility function:

$$\text{for all } x_i, x_j \subset X\text{: if } x_i \succcurlyeq x_j, \text{ then } u(x_i) \succ u(x_j),$$

where $u()$ is the utility function.

Notice that the above formula does not specify value of the utility function. In fact there is an infinite number of functions which can represent the same rational preferences of the individual[3]. So what is the utility of a loaf of bread? One function will assign it value 3, other – π. Every function describing the rational preferences of a consumer is equally good.

Notice that utility based on preferences does not have to be a measure of satisfaction. It is simply based on preferences ignoring the motives of consumer choice.

Among numerous utility functions representing consumer choices we can now take one and check the utility values assigned for the alternatives belonging to the set X. A consumer wants to maximize his total utility within the budget constraint. Then he tries to find such a bundle of goods which will bring him the highest possible value of total utility.

The solution of this problem is a point at which a quotient of marginal utility of all the goods consumed and their prices are equal, that is for every x_i and x_j the following holds (Varian 1992):

$$\frac{\frac{\partial u}{\partial x_i}}{p_i} = \frac{\frac{\partial u}{\partial x_j}}{p_j} \qquad (1)$$

where p_i is the price of a commodity on the market.

To understand this equation let us assume that a consumer chooses a point which is not optimal, at which the quotient of marginal utility of x_i to its price is lower than this value calculated for x_j. Let the prices of both goods be equal. The consumer can sell one unit of x_j and buy one unit of x_i. These changes will bring him more utility as long as the marginal utility of x_j is higher than x_i. The important thing is that the optimal bundle

[2] There are also other assumptions concerning preferences e.g. continuity, monotonicity, or convexity. They imply certain properties of utility function. However, if the set X is finite, then no other assumptions are needed to find a utility function for rational preferences.

[3] In fact, very positive transformation of utility function is also utility function.

of goods does not depend on the choice of utility function. Each function which represents the preferences of our customer will bring us the same results.

The demand of the consumer is based on the analysis of marginal utilities. The demand curve, which represents the demand value for the commodity price, can be set as following. For every price of the commodity we can find how much the price should be lowered to buy one more unit of the good. A decrease in the price of one commodity will force the rational consumer to adjust his optimal bundle of goods so to conform with formula (1). At a new optimal point the marginal utility of a commodity with the price reduced will be lower. According to the law of decreasing marginal utility this means that the consumer will buy more units of this commodity. The demand of every consumer can be summed to create the aggregate demand on the market.

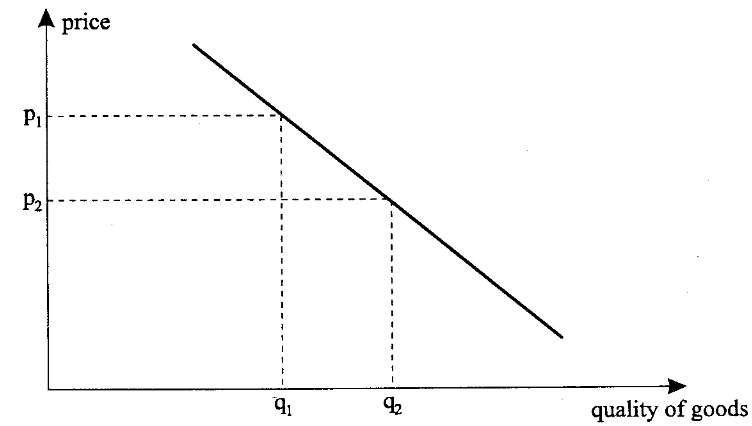


Fig. 2. The demand curve

The demand theory is based on assumptions of consumer rational choices. It turns out that these assumptions are not in reality always true. This especially concerns the transitivity property.

One example of violations of rationality postulates concerns the limited perception of a man. Imagine somebody who wants to paint his house. We show him two similar shades of yellow paint and ask which one he prefers. If the difference between them is very small then the man can be indifferent as to which one to choose. Then we show him the lighter shade of the two previous ones and a new one – a little lighter. Again the two shadows may seem so similar to the consumer that he will not notice the difference.

We can repeat the procedure until we get the colour white. Then we can show it with the darkest of all the previous ones. We expect that the man will distinguish the two paints and probably prefer one of them. In this case the preferences are not transitive.

Violation of rationality assumptions can be caused by the ways in which we present the alternatives. Kahneman and Tversky (1984) describe the following example. Imagine that you want to buy a jacket for $125 and a calculator for $15. The salesman tells you that in a store located 20 minutes away you can purchase the calculator with a discount of $5. The price of the jacket is the same there. 68% of respondents declare that they would be willing to travel to the other store to buy a calculator with the discount. If the $5 discount in the other store was only on the jacket, only 29% respondents would go to the other store. Now let's imagine that we go to the shop to buy a jacket and a calculator and we learn that because they are of stock we must go to the store 20 minutes away to buy both items. More, we learn that we get $5 on either item as a compensation. We suppose that we do not care which good will be discounted. The preferences presented in the above situations are not rational. Let's denote:

x_1 – go to the other store to purchase a calculator with $5 discount
x_2 – go to the other store to purchase a jacket with $5 discount
x_3 – buy both items in the first store

The first two situations imply that $x_1 \succ x_3$ and $x_3 \succ x_2$ whereas the second situation brings the indifference relations $x_1 \sim x_2$[4].

Another example is the changing preferences of individuals. A young smoker usually prefers to smoke one cigarette a day to smoking heavily, e.g. 20 cigarettes a day. But once he gets used to smoking he may prefer to smoke more and more. His preferences change to a point at which he prefers to smoke 20 cigarettes to one cigarette a day. If we compare the first and last choices of the smoker we conclude that they are not rational. The changing tastes of the consumer lead to violations of transitivity assumptions. The rational consumer has a stable and well defined preference order. In reality preferences may change every minute, e.g. our decisions may be influenced by emotions. This is important as in reality we do not purchase items only on payday.

[4] Such preferences are explained by mental accounting of the individuals.

2. The concept of *experienced utility*

The classic concept of utility based exclusively on the consumer's preferences has recently been questioned by Kahneman and his colleagues (see Kahneman, Wakker, and Sarin, 1997). They provide an enlightening example. Let us imagine a man who has two toasters in his kitchen. One of them is working and makes an excellent toast, whereas the other one is out of order and any time it is switched on, it gives an electric shock. Additionally, the owner of both these toasters suffers from advanced amnesia and quickly forgets his experiences. Consequently, when he comes to the kitchen to make a toast he uses, in turn, the working and the broken toaster. Now, if the economist would like to determine the utility of these two toasters for our decision-maker, based on his preferences, he would have to come to the conclusion that the man is indifferent between the toasters, and hence they have equal utility to him. We can conclude after Kahneman *et al.* (1997) that the utility concept prevailing among economists is, at least in some contexts, insufficient.

As Kahneman *et al.* (1997) bring it up, the classic concept of utility based on revealed preferences prevailed among economists due to the assumption that subjective hedonic experience i.e., experienced utility, cannot be measured. However, Kahneman and his collaborators (Fredrickson Kahneman, 1993; Kahneman, Fredrickson, Schreiber, and Redelmeier, 1993), showed that subjective feelings of an individual may be successfully measured. For example, in the experiments by Kahneman, Fredrickson, Schreiber, and Redelmeier (1993), the participants watched video clips of which one half were pleasant and the other one aversive. The participants were asked to evaluate, while watching a given clip, the intensity of the experienced pleasure or discomfort. This measure is called by the authors the *instant utility* of an episode or a *moment-based* measure of experienced utility. We can say that this is a hedonistic or affective experience measure. (The subjectively experienced intensity turns out to be an exponential function of physical values which is similar for various subjective experiences (where the exponent is specific to a given category of sensory experience).

Subsequently, after the presentation of all the clips, the participants were to evaluate *how much pleasure [or discomfort] they generally experienced while watching a given video clip*. The results obtained are considered by the authors as the measure of remembered utility. Thus, *remembered utility* is the total pleasure or displeasure reported retrospectively by the individual pleasure.

It turned out that the *remembered utility* of an episode (pleasant or unpleasant), measured as the retrospective evaluation of overall pleasure or pain, is predicted with high accuracy by the average of the most extreme instant utility recorded during the episode, and the utility recorded at the end of the episode. This was called by the authors 'The Peak and End Rule'. At the same time the retrospective evaluation of overall pleasure or pain was found not to be affected by the duration of the episode (duration neglect effect). 'The Peak and End Rule' violates the time monotonicity principle: adding moments of pain to a given adversive episode can improve the remembered utility of the episode, provided that the additional moments are less painful than the end of the original episode. We can, for instance, imagine this rule being applied by a dentist who, in order to moderate the (unpleasant) experience remembered by the patient, at the end of the appointment gives the patient an extra treatment which is also unpleasant but slightly less so than the actual end of the session.

Kahneman (2000, p. 689) has summarized these different concepts of utility as:

- *Decision utility* is inferred from observed preferences.
- *Total utility* is a *moment-based* measure of experienced utility. It is derived from measurements of moment-utility, statistically aggregated by an objective rule.
- *Remembered utility* is a *memory-based* measure of experienced utility, which is based on retrospective assessments of episodes or periods of life.

In their experiment Kahneman, *et al.* (1993) showed that remembered utility influences an individual's choice behavior. Namely, we may presume that individuals, faced with choices among different episodes, will prefer those of them which have the greatest remembered utility for them. Since remembered utilities are formed in accordance with 'The Peak and End Rule' the individual's choices should also violate the time monotonicity principle. Thus, we may expect that when making choices among various episodes (either pleasant or unpleasant) experienced previously, individuals will act in conflict with the time monotonicity condition and can go for an episode containing a period of an extra discomfort whenever the additional period ends in a less severe discomfort.

The results of an experiment carried out by Kahneman *et al.* (1993) fully supported the above presumption. Namely, they asked participants to put their hands into very cold water for several seconds. Subsequently, they informed each individual that three more tests of that type were still ahead of them. In the Short trial the first individual kept his hand for 60 seconds in

water of 14 degrees centigrade. In the Long trial an individual kept his hand for 60 seconds in water of 14 degrees centigrade and, subsequently, during the next 30 seconds, the water temperature was gradually increased by 1 degree centigrade, from 14 degrees centigrade up to 15 degrees centigrade. Then, after completion both trials (the order of these trials was randomized for different participants) individuals were given a choice for the third trial. He/she could choose between:

Short trial: to put his/her hand for 60 seconds into water of 14 degrees centigrade;

and

Long trial: to put his/her hand for 60 seconds into water of 14 degrees centigrade + keep his/her hand for the next 30 seconds in water of 15 degrees centigrade.

Most frequently the participants opted for the long trial, thus going for an episode containing an additional period of discomfort, i.e. choosing the dominated option. Such a result cannot be explained in terms of the classic utility concept. It requires the concept of experienced utility. Kahneman claims that if we employ only the decision-making utility (which is determined on the basis of preferences among episodes) and disregard experienced utility, experiments such as the one described above, as well as other serious research issues, are left unexplained.

In particular, the classic concept of utility does not allow us to test the assumption of rationality of the decision maker. The assumption is testable and the decision-maker may display even the most unusual preferences. Despite this, in many situations, one would just like to know whether and, if so, in what circumstances, human behavior is or is not rational. Does a consumer who, e.g., buys a product in a store at a higher price while he can buy an identical one at a lower price in a neighboring store violate in his/her preferences the postulates of rationality, or does his/her behavior result rather from the fact that, in addition to price, he/she is also guided by store prestige (he/she does not like shopping in a store for the general public)?

Until recently economists have not been interested in laboratory experiments run by psychologists. As Matthew Rabin (1996) put it, "Economists have traditionally contended that field evidence provides more insight than laboratory evidence" (Rabin, 1996, p. 64).

Contrary to this he proposes: "We should attempt to replace some of the current assumptions in economics with assumptions built from the systematic patterns of behavior identified by psychological research. Whatever the advantages and disadvantages of rigorously deriving conclusions from formally stated assumptions, our conclusions are likely to be more realistic if our assumptions have a better empirical base. Psychological research has accumulated enough general insights for us to convert these insights into tractable assumptions in formal economics" (Rabin, 1996, p. 62).

Indeed, recently a new paradigm has started to evolve in economics. This is known by the labels of *behavioral or experimental economics*. It refers to the attempts by economists to include psychological assumptions and methods in dealing with economic problems.

The laboratory controlled experiment has become a major research tool for developing behavioral economic theories. Most certainly, the concept of experienced utility is among those which are to be studied under this approach. We still have a lot to learn about how people perceive and misperceive their experienced utilities, how they adapt their preferences to changes in their lives (e.g. changes in wealth), etc.

One example may be a question discussed by Kahneman (2000, concerning studies on well-being (happiness). The assessments of this are typically obtained through self-reports. This type of interviews typically shows the so called treadmill effect – changes in life circumstances have small effects on subjective happiness. For example, in a classic study by Brickman, Coates, and Janoff-Bulman (1978), when the authors interviewed lottery winners and a control group, they found almost no difference in rated happiness between these two groups. However, subjective happiness (remembered utility) measured in such a way may be highly distorted and may dramatically differ from objective happiness measured through moment-based technique (moment-based utility). Indeed, as shown by Schwartz and Clore (1983) reported well-being on sunny days was significantly higher than reported on rainy days.

3. Decisions under risk and uncertainty: the principle of maximizing the expected utility

The term utility has also been used for decision analysis under risk and uncertainty. The first major step was Bernoulli attempt to explain why people neglect the expected value principle in their choices. Take, for example, a choice between a sure win of $10 and participating in a lottery which offers equal chances to win $5 or $15. Although the two alternatives have the same expected value of outcomes, people usually choose a sure win.

Another example is the St. Petersburg paradox. In this game a coin is tossed until the first head appears. The player gets a prize of $\$2^k$, where k denotes number of tosses. What is a sure equivalent of this game? The

expected value of winning this game is infinite but in reality people were not willing to pay for the participation in this game more than $4.

The above examples show that people do not respect the expected value of the game. Barnoulli noticed that the St. Petersburg paradox could be resolved if we assume that players maximize the expected value of logarithm of a price. The sure equivalent of participating in the St. Petersburg paradox game is then $4. This finding initiated the development of expected utility theory.

In 1947 von Neumann and Morgenstern developed an expected utility theory which was based on axioms concerning a player's behaviour. In this theory the uncertain outcomes are presented as lotteries with known probabilities and outcomes. We assume that players have a proper perception of the lottery. This means that decisions are made based on probabilities and their outcomes only. The sequence in which we present the alternatives should have no influence on the eventual choices. Probability 1 means a sure win. The player should also recognize that compound lotteries can be simplified to lotteries with the same conditions. Take for example the following lotteries. In the first one you win or lose $5 with equal chances. The second lottery has two stages. In the first stage the lottery offers 75% chance of qualifying for the second stage and 25% chance of ending the game with a win of $5. In the second stage the probability of $5 win is 1/3 and probability of losing $5 is 2/3.

If the player has a correct perception of lotteries, then we can ask him to show his preferences over lotteries belonging to a set X. The expected utility theory requires that these preferences comply with certain assumptions:

1) Completeness and transitivity. These properties are similar to those presented in the first part of this paper; here they should be defined for a set of lotteries X.
2) Continuity. This assumption means, that if there is a preference order over three lotteries, then there exists such α i $\beta \in]0,1[$, that:
$$\alpha L_1 + (1-\alpha)L_3 \prec L_2 \prec \beta L_1 + (1-\beta)L_3$$
3) Independence postulate. It says that for all $\alpha \in]0,1]$
$$L_1 \prec L_2 \Rightarrow \alpha L_1 + (1-\alpha)L_3 \prec \alpha L_2 + (1-\alpha)L_3$$

The above conditions are sufficient to set a utility function u which describes well consumer behaviour. It has the following property:
$$L_1 \prec L_2 \iff u(L_1) < u(L_2)$$

The player's aim is to maximize his expected utility calculated as Beroulli did it – as an expected value of utilities for every outcome of the lottery.

Notice that as in the case of utility presented in the first part of this article, this theory is also based on consumer preferences. Similarly, there is an infinite number of utility functions which describe the same preferences of an individual[5]. All these functions have a similar shape. This shape depends on risk aversion and is specific to each individual.

The shape of the utility function may be set in the following manner. We present the player with two options. The first one is a lottery in which he can either win prize W or end the game with nothing, with equal chances. Another alternative is a sure win of $0.5\,W$.

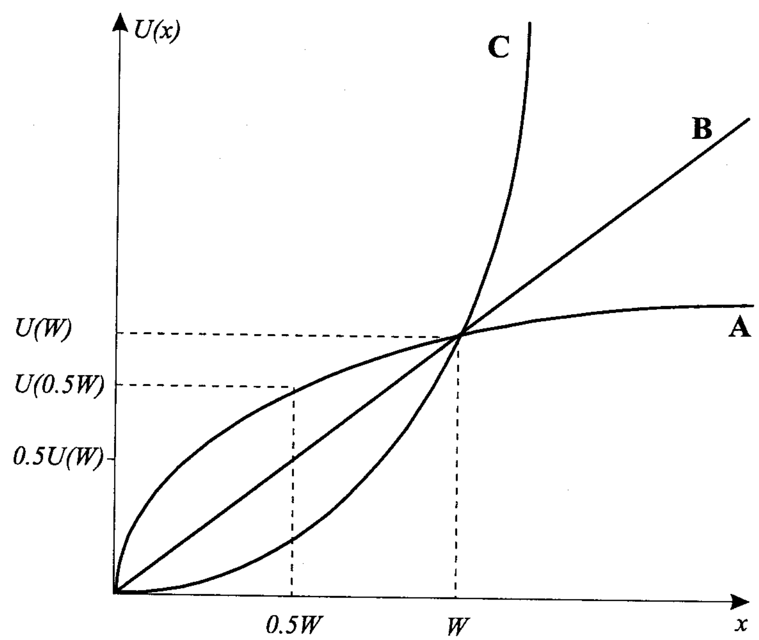


Fig. 3. Examples of utility functions

If both options are equally attractive to the player, then his risk is neutral and his utility function has a shape like the curve B in figure 3. If the player prefers a sure $0.5\,W$, then he is risk-averse and his utility function is concave like curve A. For this player the utility values follow the inequality: $u(0.5W) > 0.5u(W)$. To see how risk-averse our player is we need to find a sure equivalent of the lottery presented. The lower the value of the equivalent, the more concave the utility function is and the higher

[5] In the von Neumann-Morgenstern theory every linear positive transformation of utility function given for an individual is also the utility function.

the risk-aversion of the player. If the player is risk-seeking, then the shape of utility function is convex like the curve C. The sure equivalent is then higher that $0.5W$. An important thing is that our attitude towards risk can change as value W changes. Thus, to draw the utility function we need to find a sure equivalent for the lottery for every level of W.

The three postulates concerning preferences mean that the consumer behaviour is consistent – described by a utility function of a shape specific to an individual. In practice this assumption is sometimes violated. An example of such a choice could be the well-known Allais paradox, which is a simple comparison of preferences revealed in two situations. In the first one we present the players with two options: a sure win of \$0.5 mln or a lottery in which offers 89% probability of winning \$0.5 mln, and 10% of winning \$2.5 mln. In this lottery there is 1% chance that we end it with nothing. Most respondents prefer the first option.

The second situation is a choice between the following lotteries. The first one offers 11% chance of winning \$0.5 mln and 89% chance to end the game with nothing. The second lottery offers 10% chance of winning \$2.5 mln and 90% chance of ending the game with nothing. Most respondents choose the second option. We can show that the preferences revealed in these two situations are contradictory, in the sense of the expected utility theory – there is no utility function which describes the two choices.

The choices made in the first situation imply that:

$$u(0.5) > (0.1)u(2.5) + (0.89)u(0.5) + 0.01u(0)$$

Adding $(0.89)u(0) - (0.89)u(0.5)$ to both sides we get:

$$0.11u(0.5) + (0.89)u(0) > 0.1u(2.5) + 0.9u(0)$$

The above inequality contradicts the choice made in the second situation[6]. In this case the players violated the independence postulate.

Although the von Neumann-Morgenstern theory sometimes fails to describe people's behaviour, it is often used in economics. One of its classic applications is found in foundations of insurance theory. It provides an answer to the question: why do people buy insurance?

Let's analyse a choice made by a consumer who considers insuring his possessions, valued of W. The probability of losing them equals p. A policy which can protect him from suffering this loss costs Q. In the terms of the expected utility theory, the consumer can take part in a lottery in which he can lose W with a probability p or choose a safe loss of Q. If he decides to

[6] This paradox can be explained by "regret theory".

buy a policy then his utility would equal: $u(W-Q)$. However if he chooses to take a risk then his utility would be: $(1-p)u(W)+p\,u(0)$. If a consumer is risk averse then he is ready to pay more for a policy than the expected value of loss:

$$(1-p)u(W)+p\,u(0) < u(W-Q)$$

For the insurer the premium on the policy, apart from the expected value of losses paid out, must also cover other administrative costs and a profit margin. Risk averse consumers are likely to accept such premium. This fact enables the existence of the insurance sector in the economy.

The next example of application of the expected utility theory is from the domain of money investment. Assume that we have a certain amount of money and two types of assets. The first one is a safe bond with a constant return on investment: r_s. The other is a risky stock which is quoted on the exchange. The average and expected rate of return on this asset is r_r. According to the expected utility theory we say that even if $r_r > r_s$ the risk averse investor can allocate some money in safe bonds as it gives a sure win. In this way we can explain a demand for bonds which often have lower rate of return than stocks.

The weakness of the von Neumann-Morgenstern theory is the assumption of knowledge of all the probabilities offered by the lottery. In reality such situations occur rarely. More often we can only estimate such probabilities, for example, the chance that the fire will destroy our house in the coming year. There is no reason to think that different people will similarly asses this probability. The theory of utility based on subjective probabilities was created by Savage.

According to this theory before we take a decision, we analyse conceivable acts (decision choices) which lead to certain states of nature. The decision-maker has his own preference order over all feasible acts. He prefers the acts which result in higher expected utility of outcomes. To calculate the value of the expected utility the decision-maker uses his subjective assessment of the probabilities of outcomes. Then he compares utility of conceivable acts and sets for them a preference relation. Savage's expected utility function has the following property:

$$a_1 \succcurlyeq a_2 \iff Eu(a_1) \geq Eu(a_2)$$

where $Eu()$ is an expected value of utility;

Savage showed that the utility function exists if preferences comply with certain assumptions. The first one is the completeness of preferences over the set of alternatives. This condition is analogous to the completeness as-

sumption for the von Neumann-Morgenstern theory. The second postulate is about a sure-thing principle. It says that if we compare two acts, we base our preference relation on those states of nature which differ those acts. The third postulate means that the desirability of outcomes does not depend on the earlier acts and states – those which actually led to a present state. The fourth condition is about our subjective assessments of probabilities. The decision maker should say which events are more probable (or equally probable) for him. This means that there is a complete and transitive probability relation over all events. The last postulate was a technical monotonicity condition. Savage also formulated two additional conditions but they are not necessary to prove the existence of the utility function. The first one excluded the situation in which a decision-maker is indifferent between any two acts. The second postulate implied a sort of continuity of the preference relation and guaranteed an infinite number of all possible states.

As an example of application of Savage's theory let us consider the following example. A decision-maker has to decide whether to invest \$1000 in a venture. If he succeeds, his costs will be covered and additional profit of \$1000 earned. However, it is possible that the business will not succeed and the decision-maker will lose all the money invested. According to Savage's theory, before taking the decision he should first assess the probability of success: p. The probability of losing money is: $1-p$. In the next step the decision-maker should estimate his own utility of possible outcomes: $u(-1000)$, $u(1000)$ and $u(0)$. In the last step he chooses what decision will give him a higher value of expected utility. If:

$$u(0) > pu(-1000) + (1-p)u(1000)$$

then he will not invest his money in a venture.

Notice that in this case the decision-maker himself estimates the utility values for the outcomes. This is subjective utility which cannot be used for interpersonal comparison[7].

We can show that in some situations this theory fails to describe people's behaviour. A good example is a variation of the Ellsberg paradox (from Ellsberg 1961). We have two urns, each containing 100 balls. The balls are either white or black. It is known that the first urn contains 51 black balls and 49 white ones. The assortment of balls in the second urn is not known. One ball is picked from each urn; the colour of these balls is not disclosed. We tell a decision-maker that we can show him one of these balls, either

[7] As in the von Neumann-Morgenstern theory there is an infinite number of functions which represent preferences of the consumer. Every positive linear transformation of he utility function is also an utility function for the individual.

from the first or the second urn. If the ball disclosed is black, he gets a prize of $1000. If the ball is white, he ends the game with nothing. The majority of people want to disclose the ball from the first urn. In terms of Savage's theory it means that their subjective assessment of probability that a black ball can be picked from the second urn is lower than 51%. Then we repeat the decision problem with one modification: a prize of $1000 is given if the disclosed ball is white. It turns out that many people once again ask for a ball from the first urn. This implies that their subjective assessment of probability that white ball is picked from the second urn is lower that 49%. This contradicts the probability assessment revealed in the first step.

4. Prospect theory – critical "psychological" assumptions

There have been several proposed modifications of the utility theory for risky decision making. However, the most fundamental step in the analysis of decisions under risk was proposed in 1979 by two psychologists, Daniel Kahneman and Amos Tversky (1979). Their proposal, called the prospect theory, includes basic changes concerning the valuation of outcomes by the decision maker as well as the way it handles the probabilities. Let us start with the assumptions concerning the valuation of outcomes. They are easily presented with the aid of Figure 4.

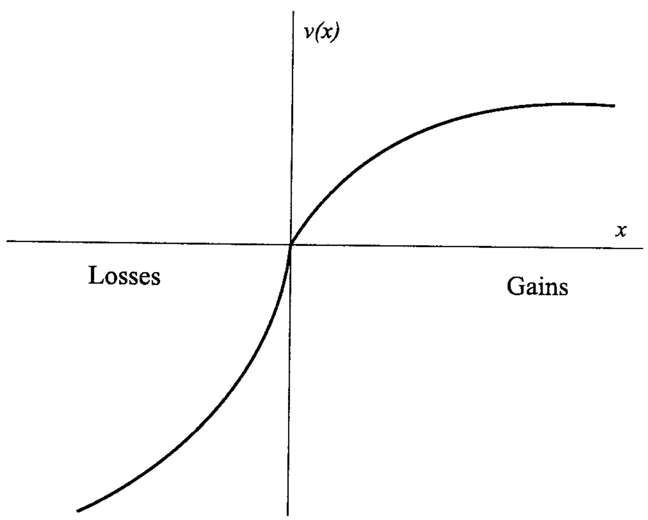


Fig. 4. Prospect theory value function

The first critical assumption is that prospects are evaluated relative to a reference point. Kahneman and Tversky (1979, p. 277) stress that people are more sensitive to how an outcome differs from some reference point than to the absolute level of the outcome itself. They put it:

"An essential feature of the present theory is that the carriers of value are changes in wealth or welfare, rather than final states. This assumption is compatible with basic principles of perception and judgment. Our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes. When we respond to attributes such as brightness, loudness, or temperature, the past and present context of experience defines an adaptation level, or reference point, and stimuli are perceived in relation to this reference point (Helson (1964)). Thus, an object at a given temperature may be experienced as hot or cold to the touch depending on the temperature to which one has adapted. The same principle applies to nonsensory attributes such as health, prestige, and wealth. The same level of wealth, for example, may imply abject poverty for one person and great riches for another-depending on their current assets."

The second assumption of the prospect theory is that preferences depend on how a problem is framed, in terms of gains or in terms of losses. As implied by the value function in Figure 1, people are risk-aversive in the domain of gains (they prefer a sure gain over a higher but risky prospect) and are risk-seeking in the domain of losses (they prefer a higher risky loss over smaller but sure loss). Tversky and Kahneman (1986) showed in a series of experiments how framing can alternate the decision-maker's choices between two logically equivalent options. One of their examples are the following decision problems.

Problem 1
Assume yourself richer by $300 than you are today. You have to choose between
a sure gain of $100
50% chance to gain $200 and 50% chance to gain nothing

Problem 2
Assume yourself richer by $500 than you are today. You have to choose between
a sure loss of $100
50% chance to lose nothing and 50% chance to lose $200

Majority of respondents preferred sure gain of $100 over fifty-fifty chance of gaining $200 or nothing in Problem 1, and in Problem 2 majority of respondents preferred fifty-fifty chance of loosing nothing or loosing $200 over sure loss of $100. However, the two problems are essentially identical. In both cases the decision maker faces a choice between $400 gain or even chance of $500 or $300. The observed framing effect results from a change in reference point.

Finally, in accordance with Figure 1, the third assumption of the prospect theory is that people dislike losses significantly more than they like gains - i.e. the hurt of a $100 loss is more noticeable than the benefit of a $100 gain. This predisposition involves loss aversion. Loss aversion has numerous implications. One is the endowment effect which involves that once a person has acquired something, he/she immediately values it more than before he/she possessed it. A simple illustration of this phenomenon is an experiment conducted by Kahneman, Knetsch, and Thaler (1990). In a classroom certain number of mugs (retail value of about $5) were placed in front of some students. Then, these students were given the following instruction: "You now own the object in your possession. You have the option of selling it if a price, which will be determined later, is acceptable to you. For each of the possible prices below indicate whether you wish to: (*x*) Sell your object and receive this price; (*y*) Keep your object and take it home with you...." These students indicated their decision for prices ranging from $0.50 to $9.50 in steps of 50 cents. Let us call them sellers. At the same time, another group of students who had not received a mug (the "choosers") were asked to indicate their preference between a mug and various sums of money ranging from $0.50 to $9.50. The authors found that the median value of the mug was $7.12 for the sellers and only $3.12 for the choosers. The difference between these values evidently reflects an endowment effect. Those students who acquired mugs valued them more than those who did not possess them.

Several phenomena in various areas of economics can be explained within the prospect theory. For example, a well known practice in pricing strategy is that price differences for the same product are typically presented as discount from the higher price rather than as premiums over the lower price. Indeed, a tourist looking for an apartment in a tourist site is typically informed that in season one pays a normal price, while out of season one pays a discounted price. Obviously, an opposite convention is equally possible that it is out of season when one pays the normal price, while in season one pays premium. But owners of the apartments know very well that the former convention introduces a more favorable reference point. The potential customer is never on the loosing side.

In the financial market Odean (1998) showed the so-called disposition effect – the tendency to hold losers and to sell winners. For example, imagine an investor who purchased two stocks A and B at $20 per share. Let us say that by the next month, stock A had decreased in price by $10 (i.e. its price has fallen to $10) and stock B had increased in price by $10 (i.e. its

price has risen to $30). The investor must decide whether to sell or to hold both the stocks for another period. Odean (1998) shows that investors are more likely to keep stock A and to sell stock B. The reason is that the initial purchase price is a natural reference point for the investor. Thus, loss aversion restrains the investor from selling the losing stock.

The prospect theory also differs from expected utility theory in the way it handles the probabilities attached to particular outcomes. The von Neumann-Morgenstern utility theory assumes that decision makers uses probabilities to calculate an expected utility of the risky alternatives. In contrast, prospect theory treats probabilities as "decision weights". The weighting function for probabilities is shown on Figure 2. As can be seen, small probabilities are overestimated while moderate and high probabilities are underestimated.

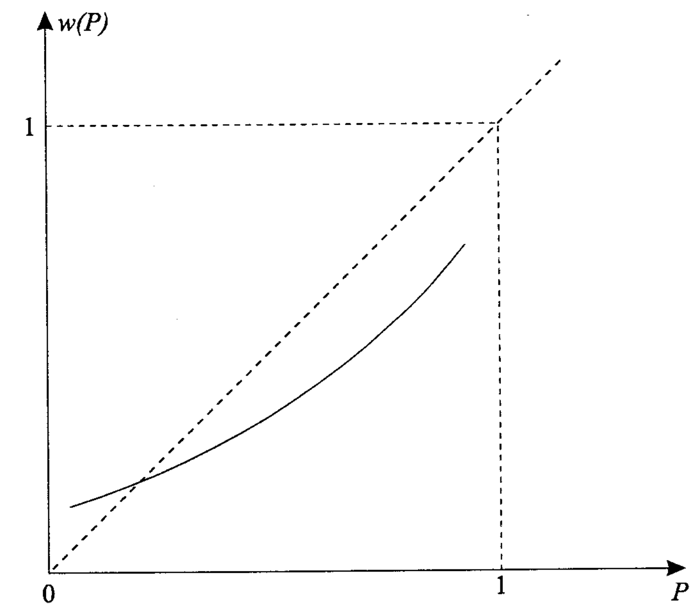


Fig. 5. Prospect theory weighting function

A simple illustration of the tendency to overweight small probabilities is the pair of problems below (cf. Kahneman and Tversky, 1979):

Problem 1.
You have to choose between:
A 1 in 1000 chance of winning $5000
A sure gain of $5.

Problem 2.
You have to choose between:
A 1 in 1000 chance of losing $5000
A sure loss of $5.

The majority of respondents who were presented with these problems preferred the risky option in the first problem and the sure loss in the second problem. Both preferences can be explained in terms of a tendency to overweight small probabilities.

The preference for the risky option in lotteries containing small probabilities of a large gain also explains why people purchase lottery tickets. The preference for the sure loss in the second problem can explain why people buy insurance. However, it is worth of noticing that in accordance with the prospect theory generally people should be reluctant to insure themselves. This is a direct inference from the assumption that people are prone to take risks in the domain of losses. Indeed, insurance can be seen as a choice between two options:

(1) to pay an insurance premium (and this way to lose some money for sure)
(2) to risk (a much) higher loss (e.g. in the case of flood) or to lose nothing (e.g. in the case of no flood).

Thus, generally, people should choose the risky option, i.e. not to insure themselves.

These ambiguous predictions of prospect theory concerning buying insurance seem to have support in real life. Kunreuther quotes a statement by George Bernstein, the former head of the Federal Insurance Administration: "Most property owners simply do not buy insurance voluntarily, regardless of the amount of the equity they have at stake. It was not until banks and other landing institutions united in requiring fire insurance from their mortgagers that most people got around purchasing it. It was also many years after its introduction that now popular homeowners' insurance caught on. At one time, insurance could not give away crime insurance, and we just need to look at our automobile insurance laws to recognize that unless we force this insurance down the throughout of the drivers, many, many thousands of people would be unprotected on the highways. People do not buy insurance voluntarily unless there is pressure on them from one source or another." (Kunreuther, 1978, p. 34).

This claim is supported by several systematic research. For instance, Brown and Hoyt (2000) report that a large portion of the property at risk from flooding in the U.S. and Germany remains uninsured. Similarily, Pynn

and Ljung (1999) found that in spite of the National Weather Service's predictions that a record flood could be produced by snow melting and the Federal Emergency Management Agency's advertisements in the media, only a minority of property owners in that area purchased insurance policies against flooding. Similarly, Tyszka et al. (2002) found that purchasing insurance against flood in the regions at risk of flooding in Poland was until the year 1997 very rare. On the other hand, immediately after the flood of 1997, when people realized that a flood is possible, the number of households insured significantly increased. However, after four further years the willingness to buy insurance in these regions started to decrease again.

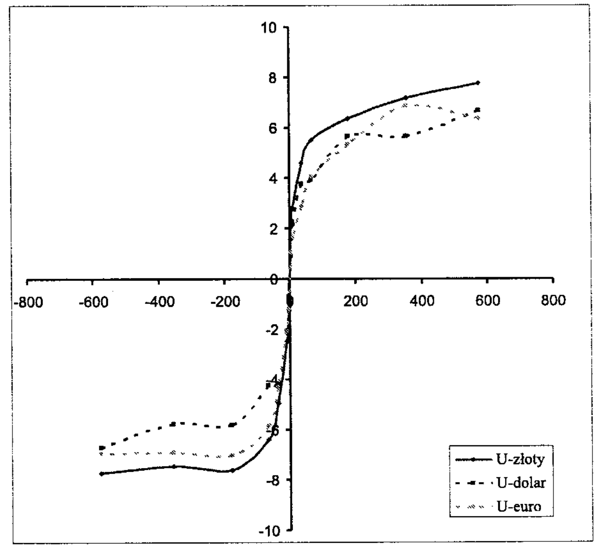


Fig. 6. Mean assessments of the importance of different gains and losses

While presenting the prospect theory, we indicated its implications in various areas of economics. To finish, let us look at its manifestation in the area of currency perception. Tyszka and Pszybyszewski (2003) asked their respondents to imagine that they gain or lose a certain amount of money or an object of a specified financial value. For different groups of respondents the values of gains and losses were expressed in different denominations: the Polish Zloty (PLN), USD and EURO. Participants were asked to express their feelings about the gains or losses, and mark them on a continuous scale ranging from "insignificant gain/loss" to "significant gain/loss". To participants who dealt with foreign currency (USD and EURO) the currency exchange rate was given so that they were able to perform a direct price comparison between foreign and PLN prices.

Figure 6 shows a direct support for the prospect theory: the value function (measured here by mean assessments of the importance of different gains and losses) was convex and relatively steep in the domain of losses. In contrast, the value function for gains was concave and less steep. At the same time, it can be also seen that both gains and losses expressed in greater nominal values were perceived by the subjects as higher. This is the so-called money illusion effect described first by Shafir, Diamond & Tversky (1997). It consists of a tendency to use the nominal value as an anchor, ignoring the real value of the money, when people evaluate the value of goods. In our case this tendency resulted in different perceptions of the same gains and losses depending on the currency they were expressed in.

References

Brickman, P., D. Coates and R. Janoff-Bulman (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality & Social Psychology* 36(8): 917–927.

Browne, M. J., and R. E. Hoyt. (2000). The Demand for Flood Insurance: Empirical Evidence. *Journal of Risk and Uncertainty*, 20, pp. 291–306.

Ellsberg, D. (1961) Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* 75: 643–669.

Fredrickson, B. L. and D. Kahneman (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality & Social Psychology* 65(1): 45–55.

Hicks, J. R., Allen G. D (1934). A reconsideration of the theory of value, cz. 1 i 2. *Economica* 1, 52–76, 196–219.

Jevons, W. S. (1871 /1970). *The theory of political economy.* Middlesex Penguin Boooks Ltd.

Kahneman, D. (2000). Experienced Utility and Objective Happiness: A Moment-Based Approach'. In D. Kahneman and A. Tversky (Eds.) *Choices, Values and Frames* New York: Cambridge University Press, pp. 673–692.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47 (March): 263–291.

Kahneman, D., B. L. Fredrickson, C. A. Schreiber and D. A. Redelmeier (1993). When more pain is preferred to less: Adding a better end. *Psychological Science* 4 (6): 401–405.

Kahneman, D., J. L. Knetsch and R. Thaler (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98: 1325–1348.

Kahneman, D., Tversky, A. (1984). Choices, values and frames. *American Psychologist* 39: 341–350.

Kahneman, D., Wakker, P. P., Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics* 112: 375–405.

Kunreuther, H. (1978). Even Noah Built an Ark. *The Wharton Magazine*: Summer, 28–35.

Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance*, 53 (5), 1775–98.

Pynn, R., and G. M. Ljung. (1999). Flood Insurance: A Survey of Grand Forks, North Dakota, Homeowners. *Applied Behavioral Science Review* 7, 171–180.

Rabin, M. (1996). Psychology and economics. *Journal of Economic Literature*, 36, 11–46.

Ricardo, D. (1957). *Zasady Ekonomii Politycznej i Opodatkowania*, Warszawa: PWN.

Savage, L. J. (1954) *The Foundations of Statistics.* New York: Wiley, 2nd. edition. 1972.

Schwartz, N. and G.L. Clore (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology* 45, 513–523.

Shafir, E., Diamond, P., Tversky, A. (1997). Money illusion. *Quarterly Journal of Economics* 112, pp. 342–374.

Smith, A. (1986/1776) *The Wealth of Nations.* London: Penguin Books.

Stankiewicz, W. (2000). *Historia myśli ekonomicznej.* Warszawa: PWE.

Tversky, A. and D. Kahneman (1986). Rational Choice and the Framing of Decisions. *Journal of Business* 59: S251–S278.

Tyszka, T., Pszybyszewski, K. (2003). Cognitive and emotional factors affecting currency perceptron. In: IAREP-workshop 2003 – proceedings. Vienna

Tyszka, T., Zaleśkiewicz, T., Domurat, A., Konieczny, R., Piskorz, Z. (2002). When are people willing to buy insurance?, *Stability and Power IAREP/SABE: Conference Proceedings*, 404–408.

Varian, H. R. (1992). *Microeconomic Analysis.* New York, London: W. W. Norton & Company.

Janusz Grzelak
Warsaw University

THE CONCEPT OF UTILITY. SHOULD IT BE REVISITED?[1]

1. Rationality as maximization of self-interest

At the beginning everything seemed simple. All rational beings are selfish, since they tend to maximize the subjective utilities of the consequences of their own actions. People are rational, thus they follow maximization rules. The rules, however, vary depending upon the type of decision making. In this paper, we will limit ourselves to decisions made under uncertainty, that is, decisions in game situations. Some implications of the paper do apply to other types of decision making as well.

In terms of game theory, any interdependence situation can be characterized by: the number of players involved, the number of actions (strategies) available to each player, the outcomes resulting from all possible combinations of actions, and each player's preferences over all outcomes. Outcomes are meant here as any action consequences that are of a (positive or negative) value to a person, be it money, other material possessions, power, status, social approval, etc. It is assumed that every individual is a rational, selfish being in the sense that he/she maximizes his/her own interest. In other words, facing a choice between two or more actions, every player is thought to choose that action whose consequences are of the highest subjective value to him/her. Since people are interdependent, what is rational to them depends upon how the individual preferences of the involved parties are related to one another: whether, and if so, to what extent and in what way, they are conflicting.

[1] Paper presented at the VI Logic, Informatics, and Philosophy of Science Workshop, Zakopane, 2003. Preparation of this publication supported by KBN grant (1 H01F 021 27), and the organizers of the Workshop.

ISBN 83-7431-015-4 ISSN 0860-150X

The basic rules of rational choice in games were first described by von Neumann, Morgenstern (1944), then developed by Nash (1951). The first rule, the maximin rule, is somewhat conservative: Choose an action the worst outcome of which is still better than the worst outcome of any other available action. This is a safe rule since it secures „the best of the worst outcomes" one can possibly obtain in a given social interaction. The second most often recommended rule to reach a solution says: Choose the action that leads to that outcome, which neither party can improve by switching to the alternative strategy. This is a simplified definition of the equilibrium solution.

The rules appear to be quite intuitive. Let us consider the abstract situations shown in Table 1 and Table 2. There are two players, A and B, each with two strategies, A and B. The numbers in the cells show the subjective value of the outcomes of each possible combination of choices for each player. The higher the number, the higher the value, i.e., the subjective expected utility.

Table 1

Prisoner's Dilemma Game (PDG)

		Column	
		A	*B*
Row	A	3 \ *3*	1 \ *4*
	B	4 \ *1*	2 \ *2*

Table 2

No Conflict Game

		Column	
		A	*B*
Row	A	4 \ *4*	2 \ *3*
	B	3 \ *2*	1 \ *1*

The logic of these two situations is different and, according to the aforementioned rules, players should choose strategy B in the PDG (Table 1) and strategy A in the No Conflict Game (Table 2). The problem, however, is that very often people do not follow these rules. At least, not when money is at stake. A review of hundreds of psychological experiments shows that a 40%, sometimes more, departure rate from theoretically rational choices is observed, even in the seemingly least problematic situations with dominant strategies, like the Prisoner's Dilemma or No Conflict Games. In other words, very often people do not choose the strategy that brings about better outcomes, regardless of what their partner(s) chooses.

2. Self-interest revisited

Does this imply that people are irrational, or should we rather reconsider the concept of self-interest used in the simplest psychological interpretations of game theory? In other words, we may agree that people satisfy their self-interest, but clarify that self-interest is complex and does not mean the same thing to all people.

2.1. Social orientations

The great diversity of people's behavior in interdependence situations can be explained only if we assume that, in an interaction with another person, people may want to achieve goals other than the mere maximization of their own profits. Although these goals are called differently in various conceptualizations and at various times (social motives, social values, social orientations) the main idea remains the same: people care not only about what they gain (or lose) themselves but also about what others gain (or lose). The subjective value of outcomes in interdependence situations is a function of both one' s own and others' outcomes (McClintock, 1972, Griesinger and Livingstone, 1973). An individual can maximize only their own outcomes (individualism), only their partner's outcomes (altruism), a joint profit (cooperative orientation), a relative gain, i.e., an advantage over their partner (competitive orientation), minimize an absolute difference between their own and their partner's outcomes (equality orientation), or seek to achieve still other goals (e.g., Wieczorkowska, 1982, Grzelak, 1982).

Figure 1 shows social orientations as a function of one's own and others' outcomes.

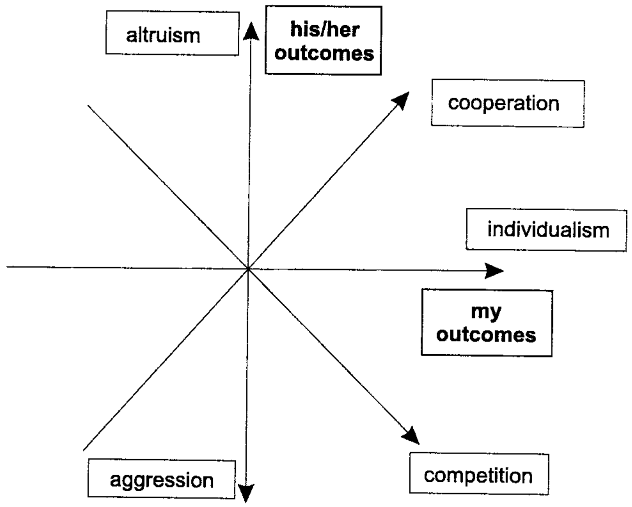


Figure 1. Social orientations in two-dimensional space of outcomes

Social orientations do not only provide a plausible *post hoc* explanation of the observed behavior. They can be measured independently of behavior and used as its predictors. A detailed description of social orientation assessment methods is beyond the scope of this chapter. However, it seems worth mentioning that some of these methods are based on a quite precise analysis of individual preferences over various allocations to the self and to the other(s). The analysis of preferences is a basis for drawing inferences about an individual's social orientation, that is, about the components of his/her subjective utilities. We assessed the social orientations of Poles in a nationwide survey (2001). Figure 2 shows the dominating orientations among Poles.

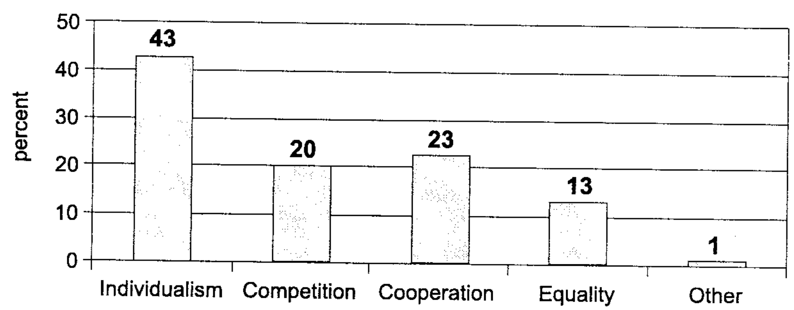


Figure 2. Dominating social orientations (2001)

Contrary to common belief, only 43% of respondents revealed a pure individualistic orientation, that is, a concern only for their own gain. The remaining 57% of population takes into account, in one way or another, others' gains or losses as well.

Although many studies have disclosed marked individual differences with respect to the goals that people strive for, social orientations are not usually considered stable, trait-like individual characteristics, since they are also situation dependent. They can vary from situation to situation depending on the partner's characteristics (wealth, prestige, identity), on group norms, and on other factors (Grzelak, 1982; 1995).

In light of these results, choosing cooperation in a Prisoner's Dilemma-like situation may be a perfectly rational decision on the condition that the persons involved in the situation care for each other's well-being. Imagine that player A values player B's outcomes with the weight of 0.6. Due to this outcome transformation, the situation is no longer a Prisoner's Dilemma. It turns out to be a rather mild conflict in which A and B's dominating strategies, A and B respectively, yield the equilibrium: A, B (see Table 3).

Table 3

Transformed Prisoner's Dilemma Game

		Column	
		A	*B*
Row	A	*3* / 3+1.8=4.8	*4* / 1+2.4=3.4
	B	*1* / 4+0.6=4.6	*2* / 2+1.2=3.2

The concept of social orientations leads to at least two important implications.

First, utility is not merely a matter of one's own gain, since it is also composed of value attached to other's gains.

Second, the weight and sign of own and others' outcomes vary from person to person. In other words, there are sound interpersonal differences in the composition of subjective utility.

Third, there is no universal rationality. Rationality is person-dependent; a question about an individual's rationality cannot be answered without knowledge of that individual's subjective utility function.

2.2. Control orientations

Any game is defined not only in terms of outcome allocations but also in terms of who decides about those allocations. Control can be considered as a means, as an ability to change outcomes in a desired direction. It can also be viewed as the end in itself.

The idea of control as a value in itself is deeply rooted in the history of psychology and it underlies a great number of psychological theories; to mention only a few: Adler's theory (1929) of power, Brehm's (1966) theory of reactance, Burger and Cooper's (1979) conception of desire for control, Winter's (1996) theory of social motivation, or Seligman's (1975) theory of learned helplessness resulting from loss of control. Although these theories are focused on different types of control motivation, they all share the same notions: (1) that people value control and (2) that the degree to which control is valued affects people's perceptions, emotions and behavior.

In interactions with others, as an interdependence actor, one can try to maximize: (1) one's control over one's own outcomes (*self-control preference*), (2) one's control over others' outcomes (*power preference*), (3) others' control over one's own outcomes (*dependence preference*), (4) others' control over others' outcomes (*respect preference*), and (5) joint control, one's and others' control over one's own and others' outcomes (*collaboration preference*) (Grzelak, 2001).

Figure 3 shows the dominant control orientations assessed in a 2001 survey run on a representative sample of adult Poles.

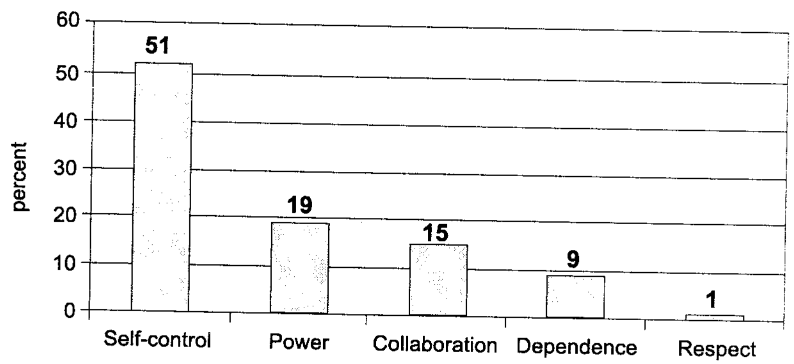


Figure 3. Dominating control orientations (2001)

It is noteworthy that only half of the population prefers to be in control over their own outcomes, and nearly a quarter prefers either passing control over their own fate to others (dependence orientation) or at least sharing it with others (collaboration). Thus, there are, as in the case of social orientations, pronounced individual differences in preferences over control.

Control preferences may affect the course of any given interaction. For instance, depending upon the type of preferences for control that are dominant, an individual either takes his/her own initiative or makes room for a partner's initiative in decision making. Turn taking in social interaction may be instrumental not only for the final outcome allocation (Kelley, 1997) but also for control allocation itself.

I began this chapter with the statement that every individual is thought to choose that action whose consequences are of the highest subjective value to him/her. That statement refers to choice between situations as well: Every individual is thought to choose that situation which is most attractive to him/her both in terms of available outcomes and in terms of control. Any social situation is evaluated, among other criteria, on the extent to which it satisfies one's control preferences. It is claimed that control preferences are one of the crucial determinants of inter-situational mobility of individuals.

The two types of orientations are theoretically and empirically orthogonal. Each of them is an independent source of attractiveness, that is, a separate component of the subjective utility of alternatives, be it various social situations or strategies available in a given situation. A large number of psychological and sociological studies have shown that people's actual behavior, emotions and cognitions are better predicted if we assume subjective utility to result from preferences over control and outcome allocation to the self and to the other(s).

3. Concluding remarks

Von Neumann and Morgenstern were not mistaken. In essence, they could not be. People do what they like. So, people maximize their subjective expected utility. For decades, researchers and theorists have assumed a simplistic interpretation, that utility is a function of one variable only: individual gain or loss. The closer the decisions were to economic ones, the easier it was for that gain or loss to be money, or other goods or services convertible to money. Even for psychologists, this vision of utility remained alluring for a long time.

The concept of utility does not need to be revisited. People still do as they like, as they did 60 years ago. What should be revisited is our interpretation of what they like. In games, a decision maker is both affected by and affects the other party. The other party is an intelligent, i.e. rational, player. This very nature of interdependence makes the distribution of control over game outcomes a non-neutral issue. In the case of people, the control issue is not neutral also because of the value assigned by the players to the game outcomes. Just how (how much and in which way) this neutrality is infringed varies from person to person, from player to player. The same situation can, then, be differently defined by different players.

The concept of utility is growing ever more complex. Its explanatory value has not been and is not under debate. Its predictive value depends on whether we can measure its most significant components, and those include the social and control orientations outlined in this chapter. In recent decades, we have observed substantial progress in the measurement of these variables. The concept of utility is growing more complex, but the measurement possibilities of contemporary psychology appear to match its level of complexity.

Bibliography

Adler, A. (1929). *The science of living.* Greenberg.

Brehm, J. V. (1972). *Responses to loss of freedom: A theory of psychological reactance.* Morristown: General Learning Press.

Burger J. M., Cooper H. M. (1979). The desirability of control. *Motivation and emotions*, 3, 381–393.

De Charms, R. (1968). *Personal causation: The internal affective determinants of behavior.* New York: Academic Press.

Griesinger, D. W., Livingstone, J. W. (1973). Toward a model of interpersonal motivation in experimental games. *Behavioral Science*, 18, 73–78.

Grzelak, J. L. (1982). Preferences and cognitive processes in social interdependence situations. In V. Derlega, J. Grzelak (Eds.). *Cooperation and helping behavior.* Theory and research. New York: Academic Press.

Grzelak, J. Ł. (1995). Sytuacyjna zmienność preferencji społecznych. Kolokwia Psychologiczne, Tom 5, *W kręgu teorii czynności.* Warszawa: IPPAN.

Grzelak J. Ł. (2001). Control preferences [w:] J. Bargh, D. Apsley (red.) *Unraveling the complexities of social life.* Washington: APA.

Kelley, H. H., Thibaut, W. (1978). *Interpersonal relations. A theory of interdependence.* New York: Wiley.

Kelley, H. H. (1997). Expanding the analysis of social orientations by reference to the sequential-temporal structure of situations. *European-Journal-of-Social-Psycholog*, 27, 4 373–404.

McClintock, C. G. (1972). Social motivation – a set of propositions. *Behavioral Science*, 17, 458–454.

Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics*, 54, 286–295.

Seligman, M. E. P. (1975) *Helplessness: On Depression. Development and Death.* San Francisco: W. H. Freeman.

von Neumann, J., Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton: Princeton University Press.

Wieczorkowska, G. (1982). A formal analysis of preferences. *Polish Psychological Bulletin*, 13, 73–77.

Winter, D. G. (1996). *Personality*, McGraw Hill.

Janusz Grzelak
Faculty of Psychology
Institute for Social Studies
Warsaw University

Witold Marciszewski
University of Białystok

POLISH NOTATION, WELL-ORDERING, AND PRAXEOLOGY

Introduction

When looking for a formal model of human action, apt to be rendered in an algorithmic way, one may find a promising possibility in a generalization of syntactic principles underlying the Polish notation. The present paper drafts such a study.[1] That syntactic theory can shed light on more general problems of intelligent behaviour as investigated in cognitive science. Moreover, logicians have their own motivation to trace connexions between syntax and general laws of thought and action.

To wit, Cantor's well-ordering theorem, to the effect that **it is always possible to arrange every well-defined set in the form of well--ordered set**, was regarded by him as a *fundamental law of thought, rich in consequences and particularily remarkable for its general validity* – Cantor [1883] (p. 169, commented by Hallett [1984], p. 73, 155 ff, etc.). The same view was expressed by Zermelo and, in a sense, by Hilbert. There are some questions as to interpretation of this principle, e.g., what the phrase "well-defined" should mean, but there is a fairly common agreement as to its validity in the realm of mathematics.

What about its validity in the empirical world? In any finite domain of physical entities they can be well-ordered in a spatial sequence. It seems also reasonable to look for other empirical domains in which the well-ordering principle could be applied. Among those worth considering there are collections of well-formed strings of a language.

---

[1] This contribution was supported with financial means of the State [Polish] Committee for Scientific Research as research project no. 2 H01A 030 25: *Nierozstrzygalność i algorytmiczna niedostępność w naukach społecznych* (Undecidability and Algorithmic Intractability in Social Sciences), planned for 2003–2006.

ISBN 83-7431-015-4 ISSN 0860-150X

Obviously, Cantor's principle holds for such collections, in particular that defined by the grammar governing Polih notation; its principles are developed in what nowadays is known as a version of categorial grammar, going back to Kazimierz Ajdukiewicz [1935] (cp. comments by Buszkowski [1996]). As far as other grammars are concerned, the problem deserves a separate research. What is examined in this study, it is a praxeological generalization of that application of the ordering principle which holds for Łukasiewicz notation, as introduced by him in *Elements of Mathematical Logic* [1929].

1. A programme for syntax of action

Polish notation has found fruitful applications in computer science, especially its versions termed as reverse Polish notation (used in machine codes and such programs as PosScript), and Cambridge Polish notation (used in programming languages of higher order as LISP); the latter reintroduces parentheses without losing, though, the syntactic principle characteristic of this notation. When addressed to computers, a string of symbols in Polish notation does not express any truth about the world. Instead, it forms a sequence of imperatives to control machine's work.

These applications are motivated by greater simplicity (from computer's point of view) of parentheses-free strings in machine codes, and are due to the main syntactic idea underlying the notation. It is the idea of a hierachical sequence of operations, i.e. functions in the mathematical sense, being materialized in machine procedures. This hierarchy is what makes it possible to discern syntactic constituents without segmenting with parentheses, provided that the number and order of arguments of each functor (i.e. function sign) is previously defined in the vocabulary of the language in question. If the number of arguments is not given as, e.g., in LISP, then the language in question combines the use of parentheses with the syntactic order of Polish notation (Cp. Robinson [1979]).

The behaviour which is controlled by instructions recorded in a machine code provides us with a fitting model to be applied in cognitive science and biology. Human behaviour can be considered as realizing a program, in particular a program for planning a future action.

There is a widely known case of Koehler's chimpanzee Sultan who fitted a bamboo stick into another, after some futile attempts to solve the problem of grasping fruit that was out of his reach. In spite of the whole distance between a human and an ape, a human would react in a similar way, as in the case in question this is the only correct solution; therefore when

reconstructing the structure of that action we are entitled to make use of our human thought-experiment. (Cp. Marciszewski [1994], p. 143 ff).

The planning proceeds from more general to more concrete statements concerning what should be done; the latter express ideas which cannot arise unless the problem is stated in general terms. In the following list of planning steps, the longer the sequence of digits, the more concrete step is numbered by it.

1. to have a tool adjusted to the distance from the agent to the fruit
1.1. to shorten the distance
1.2. to lengthen the stick (serving as the tool)
1.1.1. to come closer to the fruit
1.2.1. to join two sticks
1.2.1.1. to fasten stick no.1 to stick no.2
1.2.1.2. to fasten stick no.2 to stick no.1.

All the numbered items refer to actions being means which produce the desired result. The result is an action of higher order, e.g. the shortening of a distance which results from coming closer (by making some steps) to the object in question; the latter (1.1.1) is a concrete means to realize the former (1.1) as specified in a more general way.

Actions are arguments of the following functions:

*⇒ assigns the desired result to planned means;

n & combines two or more actions into a whole;

::: is the function which can be called sequential conjunction of more than two arguments; for instance, :::xyz means that x, y and z form a compound process, in which y follows x, and z follows y.

Thus the course of thought in Sultan's head could take the form described by the following formula (in which, for short, dots between digits are omitted).

::: [*⇒(&1211,1212)121] [&(*⇒121,12)(*⇒111,11)] [*⇒(&11,12)1]

This is to mean the following. The combination of actions 1211 and 1212 which results in 121 (that is, joining two sticks) is followed by the situation in which the fact that both old sticks are joined results in making a new stick, which is suitably longer. This, in turn, is followed by the situation in which both previous results make the agent equipped with the tool appropriate to the task in question; hence, again, the function denoted by the asterisked arrow assigns the desired result (to have a tool adjusted to the task) to the means being successively obtained, due to the actions having been undertaken.

Let it be noted that parentheses and spaces have been here added for the following two reasons: (i) for a greater perspicuity, (ii) as a provisional device which is required only then when the language in question is not sufficiently defined (e.g., as being created *ad hoc*, just to provide an example). In such a case, we need parentheses in order to indicate how many arguments should occur within the scope of the given functor, since this is not defined in an available vocabulary, as is defined, for instance, in the language of propositional calculus, arithmetic, etc. Obviously, in such a case the term 'Polish notation' ceases to be equivalent with the term 'parentheses-free notation', but the original purity of Łukasiewicz's idea can be restored, provided that kind of vocabulary in which the number of arguments following each functor is defined, and firmly fixed.

This is just an example of the syntax of planning actions. It represents only several functions, whereas there may be much more of them. A systematic study of such grammars should lead to a theory of action, or praxeology (in the sense given this term by Ludwig von Mises [1949] and Tadeusz Kotarbiński [1948], [1995]). The constraints imposed by syntactical principles of Polish notation should grant planning at a high level of preciseness because of the necessity of defining all the functions involved, and putting each of them into an exactly defined place in the string.

2. "That" – a test for Polish notation and a device in planning

When recommending Polish notation for extra-linguistic applications, the author is bound to test it in a linguistic area more challenging in this respect than are languages of logic and mathematics. There is in natural languages a syntactic construction which appears as headache if one tries to handle it with the Polish notation grammar, namely reported speech.

Let the sign "†" represent the particle "that" as involved in the string nv†s, where "n" represents a name, "v" a verb from the class containing "says", "believes", "knows", etc., and "s" hints at a sentence. There is a combinatory method to obtain various syntactic structures, and among them the following (quotation marks omitted for perspicuity).

[1] †(vn)s ... † links the sentences vn and s, functioning like a connective.

[2] (†v)ns ... † makes a new functor out of v, and this functor, in turn, combines n and s into a sentence.

[3] vn(†s) .... makes a name out of s, while v acts as functor to combine the names n and † s into a sentence.

Each of these interpretations is defended by some authors (as discussed by Marciszewski [1988]). In fact, each can be accepted as rendering one of possible intentions of the reported speech. Let us discuss the following statement in the Bible (Gen. II, 18).

The Lord God said, It is not good that the man should be alone.

Obviously, there is implicit † after "said". This may mean that the author reporting God's saying treats "said" as abbreviation of "says that", hence "that" is a part of "says that"; there are reasons to interpret "says" as having category s/n, while † transforms it into category s/ns. This exemplifies form [2].

Under another possible interpretation, † fits into category n/s since the phrase "that the man should be alone" is the name of the situation of which it is predicated "is not good"; this is represented by schema [3]. Parsing [1] may be less convincing for a logician, nevertheless it is treated by linguists as one which renders the function of "that" as a connective to combine two sentences.

Interpretation [3] hints at the reported speech as a device in planning which is characteristic of human intelligence. In planning a future course of action, a human being creates a scenario, a story which he or she tells himself/herself. Such a story may start form the phrase "Let us imagine that". Here "that" means something like "the following", and stands for the process being imagined. Thus it can be treated as making the name of such a process, in accordance with parsing [3].

Interestingly enough, it is that syntactic function of "that" which the researcher of intelligence William H. Calvin [1994] distinguishes as being specific for human cleverness. He means the capability of embedding a sentence into another sentence, like that nesting done in a rhyme for children, namely:

This is the farmer sowing the corn

That kept the cock that crowed in the morn.

Here "that" forms the name of an entity described by the sentence being nested. In a similar way we name imagined future events in the course of planning. This ability of nesting, according to Calvin, is what clearly distinguishes a human language from what one may call animal languages. More on this subject is said in the next Section which should sum up the present discussion.

3. Syntax as a constitutive factor of human intelligence

It was stated above that the grammar underlying Polish notation can improve planning of actions, and so make them more likely to succeed. Since successful (in a long run) activity is a proof of intelligence, the above statement accords with the tenets, appearing in cognitive science, that intelligence derives from human syntactic abilities. This point is expressed, e.g., in the following statement found in Calvin [1994].

> Language is the most defining feature of human intelligence: without syntax – the orderly arrangement of verbal ideas – we would be little more clever than a chimpanzee. Something very close to verbal syntax also seems to contribute to another outstanding feature of human intelligence, the ability to plan ahead.

Calvin does not define the relation of "closeness" between verbal syntax and the ability to plan ahead. Neither goes he deeper into the very notion of syntax. The present essay hints at a method of substantiating such a theory through a more elaborated statement of grammar and checking it both against specimens of planning actions and against more involved structures of natural language. This is specially worth trying in the face of so far--reaching Calvin's claims as the following.

> To understand why humans are so intelligent, we need to understand how our ancestors remodeled the ape symbolic repertoire and enhanced it by inventing syntax. [...] Human planning abilities may stem from our talent for building syntactical, string-based conceptual structures larger than sentences. [...] In this way, syntax raises intelligence to a new level. By borrowing the mental structures for syntax to judge other combinations of possible actions, we can extend our planning abilities and our intelligence.

However, as in the old riddle whether the egg or the hen was earlier, the reverse may be true, to wit that it was human intelligence that has brought syntactically organized language. There are good reasons to share Calvin's view as to the similarity of syntactic operations to planning and other intelligent actions. As to the priority of syntax, from which other performances of intelligence would derive, the question cannot be answered until the programme for the syntax of actions, as sketched above, becomes a full-fledged research project.

Anyway, if there is any grammar at all whose principles approximate those of planning structures in intelligent actions, it should take advantage of the concepts of operation and its operands, or functor and its arguments. Such constituents form an order which reflects the order constituted by the relation of means to ends; namely, the operand corresponds to a means while the whole formed out of it by the operator is what corresponds to an intended result.

The merit of Polish notation consists in showing that such a partial order can be transformed into a linear order, thus making our active lives more similar to the Cantorian ideal of well-ordering.

References

Kazimierz Ajdukiewicz [1935], "Die syntaktische Konnexität", *Studia Philosophica. Commentarii Societatis Philosophicae Polonorum.* Vol. II, 1–27.

W. Buszkowski, W. Marciszewski, J. van Benthem (Eds.) [1988], *Categorial Grammar*, John Benjamins, Amsterdam.

Wojciech Buszkowski [1995], "Grammatical structures and logical deductions", *Logic and Logical Philosophy*, No. 3, Nicholas Copernicus University Press, Toruń.

William H. Calvin [1994], "The Emergence of Intelligence", *Scientific American*, 271(4):100-107, October.

Georg Cantor [1883], "Über unendliche, lineare Punktmannigfaltigkeiten, 5", *Mathematische Annalen* **21**, 545–86.

Michael Hallett [1984], *Cantorian Set Theory and Limitations of Size*, Clarendon Press, Oxford.

Tadeusz Kotarbiński [1948], "Principes de bon travail", *Studia Philosophica. Commentarii Societatis Philosophicae Polonorum.* Vol. III, Curantibus K. Ajdukiewicz, R. Ingarden, K. Twardowski. Cracoviae et Posnaniae. 177–202 (submitted April 5, 1946).

Tadeusz Kotarbiński [1955], *Traktat o dobrej robocie* ['A Treatise on Good Work'; there exists a French edition], PWN, Warszawa.

Jan Łukasiewicz [1929], *Elementy logiki matematycznej*, Warsaw University [litographed; printed publication in 1958, Warszawa, PWN]. English translation (by O. Wojtasiewicz): *Elements of Mathematical Logic*, Pergamon Press, Oxford.

Witold Marciszewski [1988], "How freely can categories be assigned to expressions of natural language?", in Buszkowski et al. (Eds.) [1988].

Witold Marciszewski [1994], *Logic from a Rhetorical Point of View*, Walter de Gruyter, Berlin, etc.

J. A. Robinson, *Logic: Form and Function. The Mechanization of Deductive Reasoning*, University Press, Edingburgh.

Ludvig von Mises, *Human Action. A Treatise on Economics*, Yale University Press. Revised editions in 1963, 1966.

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 7 (20) 2004

Katarzyna Budzyńska
University of Cardinal Stefan Wyszyński
in Warsaw

ARGUMENTATION FROM SEMANTIC AND PRAGMATIC PERSPECTIVE

Introduction

The aim of this article is to describe some of argumentation attributes considered on the ground of logic. I will advocate that argumentation is a kind of reasoning which: (1) is based on specific inference ground, (2) is always accompanied by pragmatic elements occurring with reasoning.

I claim that argumentation has to be studied on two independent levels – objective level of *truth* separated from our knowledge, and subjective level of human *beliefs* on the truth. However, we can be interested in the properties that belong to both of these levels. We have to remember that they act independent from each other, so that argument attributes from one level may not influence attributes from another level at all. That is why I will consistently separate and highlight them in order to insure the clarity to our considerations until I develop exact tools, which will allow me to analyse actual everyday discussions on both levels simultaneously.

In first part of the article I would like to present basic notions necessary to create theory in second chapter. I will compare two levels: (i) objective level with notions of theory, formulae generally valid, semantic entailment and deduction, and (ii) subjective level with the notions corresponding with the previous ones – set of believes, formulae believed as generally valid (*topoi*), pragmatic entailment and reasoning based on *topoi*. In second part of the article I will introduce my definition of argumentation. Then I will analyse such understood discussion from two perspectives: objective – when argumentation allow to find the truth (i.e. leads disputants to true conclusions) and subjective – when argumentation is efficient (i.e. leads to persuade the audience of discussion). To present argumentation characteristics on the objective level I will apply the notice of statistical probability and on the subjective level – the notice of psychological probability.

ISBN 83-7431-015-4 ISSN 0860-150X

I. Basic notions

Our considerations will be related to the language J in which argumentations are formulated. J is a fragment of natural language and a set of sentences includes norms, values, questions, orders and decisions. Let WS_J be a set of J-sentences (both open and closed). In WS_J we will distinguish the set $\mathcal{F}$:

$$\mathcal{F} = \{\varphi : \varphi =: \text{'if } A \text{ then } B\text{', where } A, B \in WS_J \text{ and in } A, B \text{ occur } var_1, \ldots, var_n \text{ as free variables of some syntactic category}\}$$

We will use a symbolic representation as follows:

- $A, B, \alpha, \beta, A_i, B_j, \alpha_k$, where $i, j, k \in N \setminus \{0\}$ – variables representing constituents of the set WS_J,
- φ – variable representing constituents of the set F,
- L – variable representing groups of language users (the group may contain only one person),
- L_P, L_O, L_A – constants representing groups of disputants:
  L_P - proponent of argumentation,
  L_O - opponent of argumentation,
  L_A – audience of argumentation,
- S, H, As – pragmatic predicates:
  S – "assume that",
  H - "allow that",
  As – "assert that", "be sure that",
- B - predicate: "belief (in some degree) that".

1. Theory and the set of beliefs

Objective level: *Theory* T formulated in language J will be following ordered pair:

$$T = \langle \mathcal{A}, \models \rangle, \text{ where } \mathcal{A} \neq \emptyset,$$

in which $\models$ is the relation of inference on the ground of T-theory. The set A is the set consisting of logical axioms AL and specific axioms AT.

Subjective level: Before we formulate a definition of the set of beliefs, we need to specify the notion of strict believing:

(B1) $LB^*\alpha \iff$ a group of language users L ***strictly believes*** that the sentence α is true or right.

The language users believe that the sentence is right if the sentence expresses a norm, value or decision. When the formula α is a sentence in a logical sense, then it is believed to be true.

Here, the notion of believing is based upon the notion of ***subjective (psychological) probability***. The subjective probability of the sentence α for the group L (symbolized as: $P_{sub(L)}(\alpha)$) is the degree in which the group L believes in truthfulness/rightness of α, and is represented by the value from the interval of $(0, 1)$. So now we can note a following relation:

$$LB^*\alpha \iff P_{sub(L)}(\alpha) > 0,5$$

We will distinguish two degrees of strict sentence believing: assertion ($LAs\alpha \iff P_{sub(L)}(\alpha) = 1$) and hypothetical degree ($LH\alpha \iff 0,5 < P_{sub(L)}(\alpha) < 1$). We will name the sentence α, that group of language users L believes in a strict way that it is true/right, a belief of this group. The supposition ($LS\alpha \iff P_{sub(L)}(\alpha) \leq (0, 5)$ will not be considered as strict believing, because it does not generate the set of one's beliefs. We will say that it is a nonstrict acceptance and we will symbolize it as B.

Beliefs are not only related to the group of language users, but also to the moment of time in which we consider one's beliefs [Tokarz, 1993, 157–158]. The statement "$LB^*_t\alpha$" will describe such a situation taking place at the moment t that the group L believes in truthfulness/rightness of α. In cases when we do not highlight time we will understand that it occurs at any moment.

Below are stated axioms that further specify the meaning of predicate B^*:

(B2) $LB^*\alpha \Rightarrow \neg LB^*$ (it is not the case that α)

(B3) LB^*(if α then β) $\Rightarrow (LB^*\alpha \Rightarrow LB^*\beta)$

(B4) α is intuitive logical tautology or intuitive rule of reasoning $\Rightarrow LAs\alpha$

(B5) $LB^*(\alpha$ and $\beta) \Rightarrow LB^*\alpha \wedge LB^*\beta$

(B6) $LB^*\alpha \Rightarrow \alpha$ or it is not the case that α

(B7) At any moments $t_1 \neq t_2 : (LB^*_{t_1}\alpha \wedge \neg LB^*_{t_2}\alpha) \vee (LB^*_{t_1}\alpha \wedge LB^*_{t_2}\alpha)$[1]

Basing on the axiom (B3) we obtain the rule of consequence in beliefs:

(RKB) LB^*(if α then β); $LB^*\alpha$ therefore $LB^*\beta$

[1] [Marciszewski, 1972, 98–99]. Here we consider human beliefs in an idealistic way – we assume that language users are rational. However, some of assumptions indicated by axioms may be not satisfied in the everyday life.

Basing on the notions specified above, ***the set of L-group beliefs***, symbolized as S_L, is:

$$S_L = \langle B_L, \vDash^L_{pragm} \rangle^2, \text{ where } B_L \neq \emptyset,$$

in which B_L is the set of common beliefs of language users L: $B_L = \{\alpha : \alpha \in WS_J \wedge LB^*\alpha\}$, and $\vDash^L_{pragm}$ is the pragmatic relation of inference based on the ground of B_L. The set of beliefs S_L is ordinarily called as someone's outlook or philosophy of life.

The important point to notice is that, in contrast to the set $\mathcal{A}$ in theory T, set B_L can include norms, values and decisions.

2. Formulae generally valid and topoi

Objective level: In set $\mathcal{F}$ we will distinguish three separable subsets each of which consists of:

- Formulae generally valid,
- Formulae generally invalid,
- Formulae ungenerally valid.

Definition 1.1

The open formula $\varphi(var_1, \ldots, var_n) \in \mathcal{F}$ is ***formula generally valid*** on the ground of the theory $T \iff$ '$\forall var_1 \ldots \forall var_n(\varphi)$' is true in any model of theory T.

Definition 1.2

The open formula $\varphi(var_1, \ldots, var_n) \in \mathcal{F}$ is ***formula generally invalid*** on the ground of the theory $T \iff$ '$\forall var_1 \ldots \forall var_n$ (it is not the case that φ)' is true in any model of theory T.

Definition 1.3

The open formula $\varphi(var_1, \ldots, var_n) \in \mathcal{F}$ is ***formula ungenerally valid*** on the ground of the theory $T \iff$ '$\exists var_1 \ldots \exists var_n(\varphi)$' and '$\exists var_1 \ldots \exists var_n$ (it is not the case that φ)' are true in any model of theory T.

Subjective level: With regard to some of the formulae from $\mathcal{F}$, it is either absolutely impossible or only possible for the given language user to establish if they are generally valid, invalid or ungenerally valid. Such a situation occurs when the formula refers to the complex range of reality describing relations e.g. from psychology, ethics or social and economic field.

[2] Compare this notion in: [Marciszewski, 1969, 137], [Tokarz, 1993, 157].

The problem arises when a language user is forced to use such formula in reasoning, because e.g. s/he has to solve some dilemma from the area mentioned above. Since s/he does not know objective attributes of the formula, s/he assigns them subjectively. If the person believes that given formula $\varphi \in F$ is generally valid, we will say that this formula is *topoi* for the person[3]. Such subjective assignments, which not necessarily correspond to objective features of formula, sometimes are justified e.g. when individual is forced to make decisions and take actions.

Definition 1.4

The open formula $\varphi(var_1, \ldots, var_n) \in \mathcal{F}$ is ***topoi*** on the ground of S_L $\iff LB^*(\forall var_1 \ldots \forall var_n \varphi)$

When the formula '$\forall var_1 \ldots \forall var_n \varphi$' is believed by the group L to be right, we will say that φ is ***formula generally right*** for group L.

Objective and subjective level: Besides subjective believing in general validity of *topoi*, some of them are indeed valid up to a various degree. These are the *topoi* the generalization of which are sentences in logical meaning. We will distinguish set $\tau \subseteq \mathcal{F}$ that:

$$\tau = \{\varphi : \varphi \in \mathcal{F} \text{ and '}\forall var_1 \ldots \forall var_n(\varphi)\text{' is sentence in logical meaning}\}$$

Topoi from τ can be represented by the specific value of statistical probability. The conditional φ describes events of the kind $\underline{A}$ and events of the kind $\underline{B}$ anytime when free variables $var_1, \ldots, var_n$ are replaced with closed formulae of proper syntactic category. We will symbolize the open sentence φ that describes occurrence of the event of type $\underline{B}$ caused by occurrence of the events of type $\underline{A}$, as $\varphi(\underline{B}/\underline{A})$. The set of events $\underline{A}$ will be called population and symbolized as $\underline{\mathrm{A}}$. The set of events $\underline{B}$, which occur on condition that any of the events $\underline{A}$ occurred, will be symbolized as $\underline{\mathrm{BA}}$. Relative occurrence of events $\underline{B}$ following $\underline{A}$ occurrence, is the ratio of number of events $\underline{B}$, which occurs in a given population, to the number of the population's elements [Ajdukiewicz, 1965, 292]. To introduce a definition we have to assume that occurrence of $\underline{A}$ approaches infinity and that the limit of such infinity exists.

[3] The notion of *topoi* was broadly studied by Aristotle on the ground of rhetorics [Aristotle, 1990], [Aristotle, 2001]. Even though his perspective was taken here as the starting point, my notion of *topoi* differs in some aspects from created by that philosopher.

Definition 1.5

Let $\varphi(\underline{B}/\underline{A}) \in \tau$, where φ =: 'if A then B'. For any replacement of free variables with closed formulae in φ, the formula A describes event $\underline{A}$ and formula B describes event $\underline{B}$. Let $n(\underline{A})$ denote number of $\underline{A}$ which occurred and $n(\underline{BA})$ – number of $\underline{B}$, which occurred if $\underline{A}$ occurred. ***Statistical probability*** of event $\underline{B}$ caused by event $\underline{A}$ on the ground of theory T is:

$$P_{stat(T)}(\underline{B},\underline{A}) = \lim_{n(\underline{A})\to\infty} \frac{n(\underline{BA})}{n(\underline{A})}$$ [4]

For our convenience we will speak in short that $P_{stat(T)}(\underline{B},\underline{A})$ is the probability of relation described by the open sentence $\varphi(\underline{B}/\underline{A})$.

By considering definition (1.5) we obtain:

(1.1) Let $\varphi(\underline{B}/\underline{A})$ =: 'if A then B', where for any replacement of free variables with closed formulae in φ, the antecedent A describes event $\underline{A}$ and the consequent B describes event $\underline{B}$. Then on the ground of theory T:
(i) *Topoi* $\varphi(\underline{B}/\underline{A}) \in \tau$ is generally valid $(T) \iff P_{stat(T)}(\underline{B}/\underline{A}) = 1$
(ii) *Topoi* $\varphi(\underline{B}/\underline{A}) \in \tau$ is generally invalid $(T) \iff P_{stat(T)}(\underline{B}/\underline{A}) = 0$
(iii) *Topoi* $\varphi(\underline{B}/\underline{A}) \in \tau$ is probable $(T) \iff 0 < P_{stat(T)}(\underline{B}/\underline{A}) < 1$

If the value of statistical probability of the relation described by probable *topoi* is close to 1 (approaching value 1), we will say that such *topoi* is highly probable [Ajdukiewicz, 1965, 336], [Luszniewicz, 1994, 23]. Such *topoi* describe statistical relations establishing statistical laws of the theory T [Szaniawski, 1994, 18–28]. Highly probable *topoi* refer to the complex reality, in which events $\underline{B}$ are influenced not only by events $\underline{A}$, but also by accidental, unexpected ones [Luszniewicz, 1994, 21].

Definition 1.6

$\varphi(\underline{B}/\underline{A}) \in \tau$ is ***highly probable*** $(T) \iff P\{1-P_{stat(T)}(\underline{B},\underline{A}) < \varepsilon\} = 1$, where ε is any small number, $\varepsilon > 0$ and P is classical probability.

3. Pragmatic entailment based on *topoi*

Let $Roz(J)$ be the set of reasonings formulated in language J, S_J – the set of closed sentences of the set WS_J and Fin – the set of finite sets.

Definition 1.7

$(\beta_k) \in Roz(J) \iff \{\beta_1, \beta_2, \dots, \beta_k\} \in Fin$, $\{\beta_1, \beta_2, \dots, \beta_k\} \subset S_J$ and β_k is obtained from the previous sentences in the sequence $\beta_1, \beta_2, \dots, \beta_{k-1}$ basing on certain inference ground.

[4] [Wilczyński, 1980, 39], [Szaniawski, 1994, 27], [Mortimer, 1982, 44]

In further considerations the sentences: $\beta_1, \beta_2, \dots, \beta_{k-1}$, will be symbolized interchangeably as: $P_1, P_2, \dots, P_n$, and the last sentence β_k as: W. We will write: P, when we indicate the set of premises $P_1, \dots, P_n$.

I will introduce inference ground that is different from semantic entailment (which is considered on **objective level**)[5]: pragmatic entailment based on *topoi* (which has to be considered on **subjective level**). Let "$\{A_1, \dots, A_n\} \models^{L,\varphi}_{pragm} \beta$" mean that the set of sentences $\{A_1, \dots, A_n\}$ basing on *topoi* φ pragmatically entails sentence β on the ground of S_L, i.e. on the ground of the set of L-group beliefs.

Definition 1.8

Let $S_L = \langle B_L, \models^L_{pragm} \rangle$ and let $\varphi \in \mathcal{F}$ be *topoi* for the group L. $\{A_1, \dots, A_n\} \models^{L,\varphi}_{pragm} \beta \iff$ conditional 'if A_1 and ... and A_n then β' is obtained from *topoi* φ on the ground of set S_L.

Topoi φ stated in the definition above will be called a ground of pragmatic entailment.

Each inference makes the sentence β inherit some features from the elements of the set $X = \{A_1, \dots, A_n\}$. In the semantic entailment a sentence attribute of being true in any model of theory T is inherited, and in the pragmatic entailment – a sentence attribute of being believed as truth/right by the group L is inherited.

Let B^*_L be predicate of being strictly believed as true/right sentence by the group L. When the sentences from the set X entails, according to the group L, the sentence β and simultaneously the group L believes all the elements of X, then the group L will believe sentence β. ***The rule of inheritance of believing*** in truthfulness/rightness for pragmatic entailment can be stated as follows:

(1.2) $X \models^{L,\varphi}_{pragm} \beta \Rightarrow$ if $B^*_L(X)$ then $B^*_L(\beta)$.

In the next paragraph we will formulate ***the rule of inheritance of truthfulness***.

[5] Following Ajdukiewicz perspective I understand semantic entailment in broad sense, i.e. it may be founded not only upon logical general schemes, but also on generally valid ones, taken from other scientific theories [Ajdukiewicz, 1965, 99].

4. Pragmatic reasoning based on *topoi*

Subjective level: Let $Roz_{\varphi}(J)$ be the set of reasonings in language J based on *topoi* φ.

Definition 1.9

Let $\varphi \in \mathcal{F}$ be *topoi* on the ground of the set of beliefs S_L:

$$(\beta_k) \in Roz_{\varphi}(J) \iff (\beta_k) \in Roz(J) \text{ and } \exists(L \neq \emptyset)[\{\beta_1, \ldots, \beta_{k-1}\} \vDash^{L,\varphi}_{pragm} \beta_k]$$

The ground of pragmatical entailment (*topoi* φ) will also be called a ground of reasoning.

Following (1.2) and definitions (1.8) and (1.9): if $P_{sub(L)}(P) > 0{,}5$ and $P_{sub(L)}(\forall var_1 \ldots \forall var_n \varphi) > 0{,}5$ then $P_{sub(L)}(W) > 0{,}5$[6]. Hence in reasoning $Roz_{\varphi}(J)$ believing is inherited by its conclusion from the premises.

Objective and subjective level: Besides subjective believing, in reasonings based on *topoi* $\varphi(\underline{B}/\underline{A}) \in \tau$, objective statistical probability can be assigned to the conclusion. This is the probability, which indicates how often the conclusion, derived from true premises and believed *topoi* φ, is guaranteed to be true in any model of theory T.

Let us consider the theory T. When the event $\underline{B}$ occurs on condition that the event $\underline{A}$ occurred, then the consequent B and the antecedent A of conditional obtained from φ (B and A describe $\underline{B}$ and $\underline{A}$), both will be true in any model of T. As long as the occurrence of the event $\underline{A}$ is not followed by the event $\underline{B}$, the sentence obtained from *topoi* $\varphi(\underline{B}/\underline{A})$ will be false in any model of the theory T (A will be true and B will be false). While the occurrence of the events $\underline{B}$ is caused by the event $\underline{A}$, the consequent of conditional obtained from φ will always be true at the time, when true is antecedent of this sentence. In the reasoning based on *topoi* $\varphi(\underline{B}/\underline{A})$ the sentence obtained from the consequent of φ is the conclusion of this reasoning and the sentences obtained from the antecedent of φ are its premises.

We will now formulate ***the rule of inheritance of truthfulness***. It determines how often the conclusion inherits truthfulness from the premises in reasonings based on specific *topoi*.

(1.3) Let P be the set of premises and W be the conclusion of the reasoning based on *topoi* $\varphi(\underline{B}/\underline{A})$ on the ground of T. Let ζ be specified value from interval of (0,1). Let $\underline{W}_1$ be such an event when W is true, and $\underline{P}_1$ – such an event when each premise of the set P is true. A statistical probability of obtaining true conclusion in the reasoning $Roz_{\varphi}(J)$ on the ground of T from true premises, is:

$$\text{If } P_{stat(T)}(\underline{B}/\underline{A}) = \zeta \text{ then } P_{stat(T)}(\underline{W}_1/\underline{P}_1) = \zeta.$$

Following (1.3) and (1.1) we can specify some relations between "generality" of *topoi* φ and "deductiveness" of reasoning based on this *topoi*:

(i) If *topoi* φ is generally valid in T then the reasoning $Roz_{\varphi}(J)$ is deductive in T.

(ii) If *topoi* φ is generally invalid in T then the conclusion of reasoning $Roz_{\varphi}(J)$ never inherits in T truthfulness from its premises.

(iii) If *topoi* φ is probable in T then $Roz_{\varphi}(J)$ from true premises sometimes leads in T to false conclusions and sometimes to true ones.

(iv) If *topoi* φ is highly probable in T then $Roz_{\varphi}(J)$ almost always leads in T from true premises to true conclusions.

Here a question arises: why do we use such schemes in our thinking, which are only probable and not general? This is due to the complexity that characterizes the reality fields to which most of everyday reasonings is referred. In such cases, to achieve the highest possible probability, the reasoning person should choose highly probable *topoi*, which describe statistical relations. On the other hand, independently from statistical probability established by φ, there is subjective level of human beliefs. Thus as group L accepts scheme φ as *topoi* then each person from group L believes that any conclusion of reasoning based on this *topoi* φ and true premises will also be true (following the rule of inheritance of believing (1.2)).

Hence in case of any reasoning based on *topoi* (I classify argumentation as this type of reasoning), one has to consider two levels: (1) objective level referring to statistical probability of obtaining the true conclusion from *topoi* and true premises, and (2) subjective level referring to psychological probability of obtaining believed conclusion, as the consequence of the person believing in truthfulness of premises and *topoi*.

A Polish researcher Teresa Hołówka claims that everyday subjective generalizations are the result of incomplete perception of complex reality. It makes people create simplified representations of this reality. In a representation like that the objects are classified and various relations are determined among sets created in this way [Hołówka, 1998]. Basing on any generalization it is possible to formulate the *topoi* that can be applied as the foundation of reasoning. This is why people are able to act in such complexity. Despite of common lack of scientific knowledge concerning statistical probability, the generalizations are seldom "built" groundlessly. So

[6] Compare with [Ajdukiewicz, 1966, 194].

here we can put the question: if the scientific (statistical) methods cannot determine the value of objective probability of the relations described by *topoi* then how it is possible for people to distinguish which *topoi* are highly probable and which lead to false conclusion too often? It is not a purpose of this article to solve such a problem, nevertheless I will suggest a possible answer. It seems that people have though imperfect but still quite effective methods to determine statistical probability. Otherwise if one could assign only subjective probability to *topoi* of any objective (statistical) probability then reasonings based on some of these schemes would lead to believe false or even absurd conclusions (following the rule of inheritance of believing for pragmatic entailment (1.2)). This, as a result, would lead the person to wrong decisions and inefficient actions. However, people who reason on the ground of various *topoi* often make right decisions and effective measures. Thus, we may agree that in many circumstances individuals know at least approximate statistical probability of the relations described by *topoi*, and especially they are able to recognize the highly probable schemes. This knowledge may originate from the generation's wisdom and from "evolutional" adapting processes [Aristotle, 1996, 1143b]. Observations frequently made can be generalized into laws that in majority are statistical, but all laws are treated the same as general. Because each *topoi* is believed as generally valid then it is formulated in such a way as if it, indeed, was general. That is why one says: "Everything happens because of God's will", "Every man is jealous about their wives", "Every mother loves her children". Afterward, statistical laws are verified effectively during the life of later generations. The schemes that lead to false conclusions too often are eliminated – what can be compared to the "evolutional" adapting processes. It is possible that such processes eliminate not only inefficient patterns, but, in some sense, also individuals that tend to use such schemes. Since the person applies low probable *topoi* in her/his thinking, s/he acts inefficiently, what in effect – makes her/him badly adapted to life.

II. New definition of argumentation

In argumentation the moment of persuasion is substantial. If the sequence of sentences is to be an argumentation, the existence of ***parties to a dispute*** is necessary. We will distinguish three groups of disputants: proponent L_P which is the group of language users that persuades to his/their thesis, opponent L_O which is the group that rejects the arguments of proponent, and audience L_A which is persuaded to believe in truthfulness/rightness of proponent thesis. The necessity of such division is easy to observe in some cases from field of social discourse e.g. in law and politics arguments. In law disputes interchanging parts of proponent and opponent are played by a lawyer and a prosecutor. The audience is a judge/jury. The goal of the proponent is to persuade him/them to the thesis. The audience is not active in discussion i.e. its role is restricted to listening and bringing final verdict on speakers' opinions. The proponent does not intend to persuade his opponent. One can even say that their holding of the positions on the opposite sides from the beginning to the end of law-argumentation is essential for court trials. They aim only to convince passive side of argumentation – the jury or judge. Such types of discussion indicate that opponent and audience are different parts of argumentation even though those two sides may be represented by the same group. In literature this distinction is emphasized when considering *argumentum ad auditores* [Pszczołowski, 1974, 258].

In reference to above statements, following situations may appear in discussion:

- $L_P = L_O = L_A$ when a person tries to convince himself/herself [Perelman, 1984, 147],
- $L_O = L_A$ when an opponent is persuaded,
- $L_O \neq L_A$ when a proponent does not intend to convince his opponent, but audience, which does not participate directly in dialogue.

Someone's beliefs can be influenced in many different ways. For instance, persuasion may aim at audience emotions like in *argumentum ad baculum, ad crumenam or ad misericordiam.* I will not consider those cases as argumentation, which in turn I will understand as the sequence of sentences among which there are thesis and arguments justifying it. Let $Arg(J)$ be the set of simple argumentations in language J.

Definition 2.1

$(\beta_k) \in Arg(J) \iff \exists\varphi[(\beta_k) \in Roz_\varphi(J)]$ and $\exists(L_P \neq \emptyset)\exists(L_O \neq \emptyset)$ $\exists(L_A \neq \emptyset)$ [proponent L_P presents arguments $\beta_1, \beta_2, \ldots, \beta_{k-1}$ against the opponent L_O to convince audience L_A to believe in truthfulness/rightness of thesis β_k], where:

(i) L_P B β_k,

(ii) $\neg L_O$ B β_k,

(iii) L_P presents $\beta_1, \beta_2, \ldots, \beta_{k-1}$ that:

- L_P B (β_1 and ... and β_{k-1}) and
- $\{\beta_1, \beta_2, \ldots, \beta_{k-1}\} \models^{L_P,\varphi}_{pragm} \beta_k$,

(iv) the objective of L_P is that: $L_A B^* \beta_k$.

The set of argumentation is classified in literature according to many different criteria. These classifications aim to organize a very complicated scope of various persuasion methods. Most frequently indicated are honest and dishonest arguments (or in other words: rhetorical and eristic). Following Aristotle, I will specify the rhetorical argumentation as the one fulfilling three conditions: *logos*, *ethos* and *pathos*[7].

Definition 2.2

(β_k) is ***rhetorical argumentation*** based on $\varphi \iff (\beta_k) \in Arg(J)$ and when (β_k) fulfils following conditions:

(i) *logos*: φ is generally valid or highly probable *topoi*,

(ii) *ethos*: $P_{sub(Lp)}(\forall var_1 \ldots \forall var_n \varphi) > 0{,}5$ and $P_{sub(Lp)}(\beta_1$ and $\ldots$ and $\beta_k) > 0{,}5$,

(iii) *pathos*: argumentation (β_k) is built according to rules of stylistics.

The discussion satisfies the condition *logos* when the reasoning is deductive or leads to the false conclusion very seldom. The schemes highly probable may be selected in honest arguments only when general *topoi* are not available. This way the probability of obtaining true conclusion, i.e. $P_{stat(T)}(\underline{W}_1, \underline{P}_1)$, is the highest one can reach. Otherwise less probable *topoi* would lead to false conclusion too often and argumentation would become unreliable.

The conditions (i) and (iii) of the definition (2.1) require that proponent only expresses the elements of (β_k), but not necessarily strictly believes in truthfulness/rightness of these sentences, since it is sufficient that s/he does believe them in a nonstrict way. However, if the discussion is to be honest, it has to fulfill condition *ethos*, which means that proponent has to strictly believe sentences in (β_k). Otherwise, as the set of her/his beliefs does not contain these formulae $(\neg L_P B^* \alpha \iff \alpha \notin \mathrm{B}_{Lp})$, s/he convinces others to believe in something using premises that s/he does not believe her/himself.

Argumentation that does not fulfill at least one of conditions (i)–(iii) in definition (2.2) will be called ***eristic argumentation***. If argument does not meet condition *logos* then the inference foundation is invalid or low probable *topoi*. A proponent is either unaware that statistical probability of the described relation is too low or s/he deliberately uses "catchy", but low probable *topoi*. In second case the argumentation is unsatisfactory for not

[7] According to Aristotle, when considering any argumentation one should examine relation of semantic entailment between premises and its conclusion (*logos*), credibility of proponent (*ethos*) and mood of the audience that is influenced by stylistic methods used during the persuasion (*pathos*) [Aristotle, 2001, 1356a], [Nieznański, 2000, 118].

only *logos*, but also *ethos*. The condition *ethos* is not fulfilled when L_P just assumes (believes in supposition degree) premises, conclusion or the inference ground of given argumentation. S/he may just take into account if the audience strictly believes premises and *topoi* regardless of what proponent's beliefs really are. Anyway, it will be warranted to her/him that the audience will strictly believe the proponent's thesis (according to the rule of inheritance of believing). However believing the proponent's thesis is essential both in rhetorical and eristic arguments, it is achieved in the first type of persuasion by the fulfillment of all the conditions: *logos*, *ethos* and *pathos*, whereas in the second type – by whichever way.

Now I wish to present argumentation on two independent levels. In the first paragraph I will consider its attributes on objective level and I will attempt to determine when a discussion leads to true conclusions. In the second paragraph I intend to study argument features on subjective level. I will investigate when a discussion leads a proponent to achieve her/his main goal i.e. to persuade the audience. In the last paragraph I will describe the arguments most common in everyday life.

1. "Deductiveness" of inference schemes in argumentation

Argumentation is the reasoning based on *topoi* i.e. on the scheme, which is believed as generally valid/right. This is why *topoi* are formulated as general sentences. As a result, the pragmatic entailment seems to be the semantic entailment (the foundation of reasoning is believed as general even though actually it is not). And some of the language users may share the impression that a given argumentation is deductive.

Regardless of subjective human knowledge concerning "generality" of schemes, the formulae from set τ are objectively represented by the specific degree of this "generality". When *topoi* is a generally valid scheme then argumentation is deductive. In turn, if the discussion is based upon *topoi* generally invalid then the statistical probability of relation described by *topoi* equals 0, i.e. $P_{stat(T)}(\underline{B}, \underline{A}) = 0$. Following (1.3), we obtain that: $P_{stat(T)}(\underline{W}_1, \underline{P}_1) = 0$. Thus, in case of each argumentation based on such *topoi*, its conclusion will not inherit truthfulness from the premises.

In everyday life the most frequent arguments are those which are founded upon probable *topoi*. For such schemes $\varphi(\underline{B}/\underline{A})$ we have: $0 < P_{stat(T)}(\underline{B}, \underline{A}) < 1$. Hence from (1.3) we obtain: $0 < P_{stat(T)}(\underline{W}_1, \underline{P}_1) < 1$. If the inference ground is *topoi* highly probable even though the argumentation is not deductive and does not always lead to true conclusions, it happens almost always.

Theorem 1.

Let (β_k) be argumentation based on *topoi* φ and let $\nu(\alpha)$ represent a logical value of the sentence α. If $P_{stat(T)}(\varphi) = \zeta$ and $\nu(\beta_1) = 1$ and ... and $\nu(\beta_{k-1}) = 1$, then $P_{stat(T)}(\nu(\beta_k) = 1) = \zeta$.

The theorem (1) is the consequence of the rule of truthfulness inheritance formulated in (1.3). When argumentation is based upon *topoi* φ and true premises: $\beta_1, \ldots, \beta_{k-1}$, then statistical probability of obtaining true conclusion in this argumentation is the same as probability of relation described by scheme φ on which we "built" our argumentation. Thus, when we argue, the higher the statistical probability of relation described by *topoi* is, the higher the warranty of obtaining true conclusions is. If our **goal in discussion is objective**, i.e. we aim to "find the truth", we should select *topoi* with the highest statistical probability that is available for the subject under dispute.

There are a number of points to observe:

- As the audience does not know the degree of "deductiveness" of argumentation, the proponent may deliberately use *topoi* generally invalid or low probable taking advantage of their lack of knowledge. This kind of persuasion is called in literature *argumentum ad ignorantiam.*
- To satisfy the condition *logos*, the proponent aims to approach minimal probability of obtaining the false conclusion.
- Many arguments are based on *topoi*, which are not elements of set τ. These schemes have a following form: *if A then B*, where at least one sentence obtained from A or B expresses the norm, value or decision. Such *topoi* can be believed as formula generally right on the ground of set of L-groups beliefs. In those arguments the conclusion inherits the attribute of strict believing in rightness from its premises.

Summarizing – argumentation is a reasoning in which from true premises it is guaranteed to obtain:

(i) always true conclusions (these argumentations are deductive),
(ii) sometimes true, sometimes false conclusions,
(iii) never true conclusions or
(iv) conclusions that cannot be considered as true/false – this is when argumentation is based on *topoi* believed as generally right.

What is characteristic of the actual persuasion is that the most frequent argumentations are the reasonings of type (ii) and (iv), since in everyday life we discuss about what is uncertain or what concerns values. It differs arguments from many other reasonings, especially the ones from scientific theories.

2. The efficiency of argumentation

The second substantial difference between arguments and other reasonings is the degree to which language users participate in them. The influence of such "participation" like e.g. disagreement over the conclusion by given individuals, is not substantial in deduction at all. The only sufficient condition here for obtaining true conclusion is truthfulness of premises and semantic entailment, no matter who believes it or not[8].

We observe an opposite situation in persuasion. The reasoning is not an argumentation, when one of the following is missing: proponent (it is indicated by (i) and (iii) of definition (2.1)), opponent (ii) and audience (iv). We say that argumentation is efficient when a proponent reaches her/his goal and persuades the audience in her/his favor.

Definition 2.3

(β_k) is ***efficient simple argumentation*** for the audience $L_A \iff (\beta_k) \in Arg(J) \wedge L_A B^* \beta_k$.

To achieve efficiency of argumentation, it is neither necessary nor sufficient that (a) premises are true and (b) premises semantically entail conclusion (like in deduction), but that (1) premises are believed by audience and (2) premises pragmatically (according to audience) entail conclusion. The first two conditions (a and b) are not sufficient when the individuals do not know that the *topoi* is generally valid, so they will not be aware that argumentation based on it is deductive. As a result, it may happen that the audience will not believe the thesis that is really true. "The formal correctness" of argumentation is not the necessary condition of its efficiency either, because the audience may believe the conclusion of argumentation that is not deductive (moreover, a thesis can be false). That is why we quite often observe in everyday life that people believe as true/right the conclusions, which actually are false/wrong. So the efficiency of discussion depends more on the participation of language users than on truthfulness and its "deductiveness" [Perelman, 1984, 147], [Korolko, 1990, 40].

Theorem 2.

If $P_{sub(LA)}(\forall var_1 \ldots \forall var_n \varphi) > 0{,}5$ and $P_{sub(LA)}(\beta_1$ and ... and $\beta_{k-1}) > 0{,}5$, then the argumentation (β_k) based on *topoi* φ is efficient for the audience L_A.

[8] Perelman, for instance, says that truth is impersonal [Perelman, 1984, 147].

It is easy to show that theorem (2) is the consequence of definition (2.3) and the rule of inheritance of believing (1.2). To achieve efficiency of simple discussion with regard to the given audience, it is sufficient that the audience strictly believes inference ground φ as generally valid (i.e. as *topoi*) and premises as true/right. When following (1.2), we obtain that the group will believe the thesis of the presented argumentation and this finally, according to the definition (2.3), will mean that argumentation will become efficient. If our **goal in discussion is subjective**, i.e. we aim to persuade audience, then we should select as a scheme φ and premises the sentences (describing subject under dispute) with the highest psychological probability for the audience.

The simple argumentation is ***inefficient*** when audience finds faults in this argumentation[9]. When $L_O \neq L_A$ then the audience finds the fault either on their own or influenced by the opponent who indicates e.g. falseness of specific premise.

In argumentation that contains $P_1, \ldots, P_n$, W and is based on φ, the audience L_A does not believe W when:

(i) $\neg L_A B^* P_i$, where $i \in \{1, \ldots, n\}$, because:
- $L_A B^*$(it is not the case that P_i)
- $L_A B^*$(P_i is not well-founded)

(ii) $\neg(\{P_1, P_2, \ldots, P_n\} \models^{LA,\varphi}_{pragm} W)$, because:
- $L_A B^*$(it is not the case that $\forall var_1 \ldots \forall var_m \varphi$)
- $L_A B^*$('$\forall var_1 \ldots \forall var_m \varphi$' is not well-founded)

3. Complex argumentations

The simple argumentation is inefficient if the audience puts forward at least one of the counterplea mentioned above. However, in everyday life the persuasion may be continued. When the next reasoning is presented, the argumentation becomes complex.

In complex discussion a first simple argumentation may be followed by the next one when: either the proponent continues persuasion or the opponent presents her/his own argumentation, in which $\neg\beta_k$ is a conclusion (if a conclusion of previous argumentation was β_k).

[9] In literature it is called counterplea for the argumentation [Nieznański, 2000, 117], [Łuszczewska-Romahnowa, 1966, 164].

Thus ***the complex arguments*** can have following forms:
- arguments with invariable proponent or
- arguments with variable proponent.

A. Complex argumentations with invariable proponent

To achieve efficiency, the proponent may present a second simple argumentation in which: (1) the conclusion is the scheme φ or the premise, that audience did not believe in first argumentation, or (2) the conclusion stays unchanged, but the new argumentation is based upon other scheme or other premises.

Example 2.1

Let us assume that at the moment t_1, in which the first simple argumentation is presented, the audience believes premises of reasoning. However, it does not believe inference ground as generally valid by claiming that it is insufficiently founded:

(Arg1) $\{P_1, \ldots, P_n\} \models^{Lp,\varphi_1}_{pragm} W$,

and $L_A B^*(P_1$ and ... and $P_n)$ and $L_A B^*_{t_1}$('$\forall var_1 \ldots \forall var_a \varphi_1$' is ill-founded).

Because $L_A B^*_{t_1}$('$\forall var_1 \ldots \forall var_a \varphi_1$' is ill-founded) then $\neg L_A B^*_{t_1} W$.
Thus, at the moment t_2 a proponent presents premises $Q_1, \ldots, Q_k$, that according to him entail as conclusion the sentence: $\forall var_1 \ldots \forall var_a \varphi_1$. S/he selects now as the inference ground a new scheme: $\varphi_2(var_1, \ldots, var_b)$, in which $var_1, \ldots, var_b$ are free variables of some syntactic category.

(Arg2) $\{Q_1, \ldots, Q_k\} \models^{Lp,\varphi_2}_{pragm} \forall var_1 \ldots \forall var_a \varphi_1$.

Let us assume also that:

$$L_A B^*_{t_2}(Q_1 \text{ and ... and } Q_k) \text{ and } \{Q_1, \ldots, Q_k\} \models^{LA,\varphi_2,t_2}_{pragm} \forall var_1 \ldots \forall var_a \varphi_1.$$

Thus following the rule of inheritance of believing (1.2), we obtain: $L_A B^*_{t_2}(\forall var_1 \ldots \forall var_a \varphi_1)$, therefore: $\{P_1, \ldots, P_n\} \models^{LA,\varphi_1,t_2}_{pragm} W$. And because: $L_A B^*(P_1$ and ... and $P_n)$ then: $L_A B^*_{t_2} W$.
Hence finally at the moment t_2, the complex argumentation (Arg) (that contains (Arg1) and (Arg2)) becomes efficient, because $L_A B^*_{t_2} W$.

Example 2.2

Let us assume now that the first simple argumentation is the same as in the above example. However, this time to persuade the audience at the moment t_2, the proponent presents the second argumentation based upon the other sentence: $\varphi_3(var_1, \ldots, var_c)$, where $var_1, \ldots, var_c$ are free variables of some syntactic category:

(Arg3) $\{R_1, \ldots, R_m\} \vDash^{Lp,\varphi 3}_{pragm} W$.

Let us also assume that

$$L_A B^*_{t_2}(R_1 \text{ and } \ldots \text{ and } R_m) \text{ and } \{R_1, \ldots, R_m\} \vDash^{LA,\varphi 3,t_2}_{pragm} W.$$

Following the rule of belief inheritance (1.2) we obtain: $L_A B^*_{t_2} W$.
Thus, if the audience believes premises and inference ground of (Arg3) then the complex argumentation becomes effective in t_2, even though (Arg1), that led to the same conclusion as (Arg3), was inefficient.

B. Complex argumentations with variable proponent

Argumentation with variable proponent always includes at least one counterargumentation. Let us consider the following example:

Example 2.3

Let t_1 be the final moment of (Arg):

(Arg) $\{P_1, \ldots, P_n\} \vDash^{Lp1,\varphi_1}_{pragm} W_1$,

and $L_A B^*_{t_1} W_1$.

From the definition (2.1) we know that: $\neg L_O B^*(W_1)$. If the opponent does not want the audience to keep believing the proponent's thesis W_1 then s/he may present a counterargumentation with conclusion W_2 which is: W_2 = 'it is not the case that W_1'. The following simple argumentation with the inference foundation: $\varphi_2(var_1, \ldots, var_b)$, is now presented:

(KArg) $\{Q_1, \ldots, Q_m\} \vDash^{Lp2,\varphi_2}_{pragm} W_2$,

Hence, the opponent and the proponent of previous argumentation "changed their parts with each other". It should be noted that: $L_{P2} = L_{O1}$, $L_{O2} = L_{P1}$, $L_{A2} = L_{A1}$, where L_{P2}, L_{O2} and L_{A2} are participants of counterargumentation (KArg) and L_{P1}, L_{O1} and L_{A1} are participants of (Arg).

Let us assume now that (KArg) is efficient. Thus at its final moment t_2: $L_A B^*_{t_2}$(it is not the case that W_1). Following the axiom (B2) and rule (RKB) we obtain: $\neg L_A B^*_{t_2}$(it is not the case that it is not the case that W_1). Assuming that: $\neg\neg\alpha \Rightarrow \alpha$ is intuitive tautology and following (B4) and (RKB) we obtain that: $\neg L_A B^*_{t_2} W_1$.

In consequence at the moment t_2, argument (Arg) is inefficient, because counterargumentation (KArg) becomes efficient.

C. Efficiency of complex argumentations

Basing on the above examples we will now formulate the definition of efficiency of complex discussions:

Definition 2.4

Let complex argumentation Arg_{cplx} be presented in time-period, where t_1 is the beginning and t_j is the end of this period ($1 < j, j \in N$). Let W_{Lp} be the conclusion of simple argumentation in which L_P is proponent and W_{Lp} is not a premise or an inference ground of any other simple argumentation in Arg_{cplx}.

Complex argumentation Arg_{cplx} ***is efficient*** for the proponent L_P and audience $L_A \iff L_A B^*_{t_j} W_{Lp}$.

In the example (2.1) the complex discussion is efficient for the proponent L_P and the audience L_A, because even though $\neg L_A B^*_{t_1} W$, but $L_A B^*_{t_2} W$, and it was t_2 that was the final moment of complex argument. We do not consider the efficiency with regard to the conclusion of simple argumentation (Arg2): $\forall var_1 \ldots \forall var_a \varphi_1$, because it was the inference ground of other argumentation i.e. (Arg1). In the example (2.2) complex argumentation is efficient for L_P and L_A, because $L_A B^*_{t_2} W$, although in this case too first simple argumentation was inefficient and the audience did not believe the thesis W in the beginning, i.e. $\neg L_A B^*_{t_1} W$. In the example (2.3) complex argumentation is efficient for the proponent L_{P2} and audience L_A, because in the end (in t_2) the audience believed the conclusion of argumentation in which L_{P2} was proponent, i.e. $L_A B^*_{t_2} W_2$. While W_1 was believed by the audience, efficiency could not be compared with this sentence, because t_1 was not the final moment of complex discussion. And in moment t_2 we have: $\neg L_A B^*_{t_2} W_1$. Thus, however, the first simple argumentation was efficient for proponent L_{P1}, the whole discussion was "won" by the proponent of counterargumentation, i.e. L_{P2}.

The argument efficiency is related to the audience's set of beliefs, which in turn is related to time. In example (2.1) in the beginning the audience did not believe φ_1 as generally valid, i.e. did not believe the inference ground of the first simple argumentation (Arg1): $\neg L_A B^*_{t_1}(\forall var_1 \ldots \forall var_a \varphi_1)$. And because: $B_{L,t} = \{\alpha : LB^*_t \alpha\}$ then: '$\forall var_1 \ldots \forall var_a \varphi_1$' $\notin B_{LA,t_1}$. Thus, following the definition (1.8) on the ground of the audience's set of beliefs at the moment t_1, the premises P_1 and ... and P_n do not pragmatically entail the conclusion W: $\neg(\{P_1, \ldots, P_n\} \vDash^{LA,\varphi_1,t_1}_{pragm} W)$. However, in (Arg2) the conclusion '$\forall var_1 \ldots \forall var_a \varphi_1$' was pragmatically entailed on the ground of S_{LA,t_1}. In this way at the moment t_2 the formula '$\forall var_1 \ldots \forall var_a \varphi_1$' was added to the set of audience's beliefs i.e.: $S_{LA,t_2} = \langle B_{LA,t_1} \cup \{`\forall var_1 \ldots \forall var_a \varphi_1'\}, \vDash^{LA,t_2}_{pragm} \rangle$. So at the moment t_2 it was possible to derive the sentence W on the ground of the set of L_A-beliefs, i.e. if '$\forall var_1 \ldots \forall var_a \varphi_1$' $\in B_{LA,t_2}$, then $\{P_1, \ldots, P_n\} \vDash^{LA,\varphi_1,t_2}_{pragm} W$. And because ($P_1$ and ... and P_n) $\in B_{LA,t_2}$, then $W \in B_{LA,t_2}$.

In the above examples we assumed the simplification that the audience believes the conclusion in the second step (in the second reasoning). In everyday persuasion the discussion may be much more complex. The statements formulated above can be, of course, generalized on any long sequence of simple argumentations that we can observe in day-to-day life.

Once we become interested in the issue of social discourse, we ought to consider the two levels. As long as our main concern is a victory in dispute, we stay on the subjective level. In order to convince the audience to believe our thesis, we have to select for our argumentation such premises and inference ground, which are believed by this audience (i.e. with the highest available psychological probability for this group). While we aim to "find the truth" by discussing with someone, we are on objective level. Our only concern then should be to provide our argumentation with premises that are true and inference scheme, which is general or at least highly probable (i.e. with the highest available statistical probability). In such case we may neglect disputants beliefs, in particular if selected scheme is *topoi* for our audience or it is not. However, as our purpose is both convincing and cognition, we have to connect these levels. Thus, it is necessary that we select: (1) scheme **subjectively** believed by the audience as *topoi*, which **objectively** is generally valid/high probable, and (2) premises **subjectively** believed by the audience and which **objectively** are true. Since we are aware of the presence of these two independent levels in discussion, it makes us understand better the principles governing the argumentation and furthermore helps us to achieve both goals of high importance in everyday persuasion.

References

Ajdukiewicz K., *Logika pragmatyczna*, PWN, Warszawa 1965

Ajdukiewicz K., *Systemy aksjomatyczne z metodologicznego punktu widzenia* [in:] *Logiczna teoria nauki*, PWN, Warszawa 1966, 187–204

Aristotle, *Etyka nikomachejska* [in:] *Dzieła wszystkie tom 5*, PWN, Warszawa 1996

Aristotle, *O dowodach sofistycznych* [in:] *Dzieła wszystkie tom 1*, PWN, Warszawa 1990

Aristotle, *Retoryka* [in:] *Dzieła wszystkie tom 6*, PWN, Warszawa 2001

Aristotle, *Topiki* [in:] *Dzieła wszystkie tom 1*, PWN, Warszawa 1990

Borkowski L., *Wprowadzenie do logiki i teorii mnogości*, Towarzystwo Naukowe KUL, Lublin 1991

Hołówka T., *Błędy, spory, argumenty. Szkice z logiki stosowanej*, Wydział Filozofii i Socjologii Uniwersytetu Warszawskiego, Warszawa 1998

Korolko M., *Retoryka i erystyka dla prawników*, PWN, Warszawa 2001

Korolko M., *Sztuka retoryki. Przewodnik encyklopedyczny*, Wiedza Powszechna, Warszawa, 1990

Luszniewicz A., *Statystyka nie jest trudna. Metody wnioskowania statystycznego*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 1994

Łuszczewska-Romahnowa S., *Pewne pojęcie poprawnej inferencji i pragmatyczne pojęcie wynikania*, [in:] *Logiczna teoria nauki*, PWN, Warszawa 1966, p. 163–167

Marciszewski W., *Podstawy logicznej teorii przekonań*, PWN, Warszawa 1972

Marciszewski W., *Sztuka rozumowania w świetle logiki*, Aleph, Warszawa 1994

Mortimer H., *Logika indukcji. Wybrane problemy*, PWN, Warszawa 1982

Nieznański E., *Logika. Podstawy – język – uzasadnianie*, C. H. Beck, Warszawa 2000

Ossowska M., *Podstawy nauki o moralności*, PWN, Warszawa 1966

Patryas W., *Uznawanie zdań*, PWN, Warszawa – Poznań 1987

Perelman Ch., *Logika prawnicza. Nowa Retoryka*, PWN, Warszawa 1984

Przełęcki M., *Pojęcie prawdy w językach nauk empirycznych*, [in:] *Studia Filozoficzne* nr 6 (139), 1977, 13–20

Pszczołowski T., *Umiejętność przekonywania i dyskusji*, Wiedza Powszechna, Warszawa 1974

Szaniawski K., *O nauce, rozumowaniu i wartościach*, PWN, Warszawa 1994

Szymanek K., *Sztuka argumentacji. Słownik terminologiczny*, PWN, Warszawa 2001

Tokarz M., *Elementy pragmatyki logicznej*, PWN, Warszawa 1993

Wilczyński J., *Prawdopodobieństwo – interpretacje pojęcia*, [in:] *Studia Filozoficzne* nr 1 (170), 1980, 35–47

Wróblewski J., *Wartości a decyzja sądowa*, Zakład Narodowy im. Ossolińskich, Wrocław 1973

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 7 (20) 2004

Simon Czarnik
Jagiellonian University

VOLUNTARY AND FORCED REDISTRIBUTION UNDER DEMOCRATIC RULE

There is a wide literature on the problem of division of goods from a perspective of both social choice theory and game theory. The former scrutinizes formal properties of different allocation rules (usually referred to as competing definitions of 'distributive justice')[1], while the latter usually concentrates on the interaction of strategies employed by actors striving to achieve the best preferred division of goods in question. It should be stressed that 'best preferred' is not necessarily tantamount to 'self-interest maximizing'. Indeed, as the ample empirical evidence shows, the standard game-theoretical assumption of individual egoism does not hold true in some of laboratory games played by human subjects. In recent years we have been witnessing a rapid growth of experimental research conducted to test for alternative explanations of distributive behavior, such as *altruism or spitefulness* (Levine 1997), considerations for *fairness* or *reciprocity* (Fehr Schmidt 1999; Tyran Sausgruber 2002; Bolton Ockenfels, forthcoming), or *empathic responsiveness* (Fong 2003). On the other hand, self-interest was found to be of utmost significance in experiments where subjects' payoffs were dependent on their own effort/productivity (Rutström Williams 2000, Gächter Riedl 2002), as well as in some games with random entitlements, e.g hawk-dove game (Neugebauer Poulsen Schram 2002).

Games typically used to model an intentional division of goods are various types of dictator, ultimatum, and gangster games. They all refer to the problem of 'splitting the cake': dictator game assumes that one player, who is initially endowed with entire cake, is absolutely free to define the ultimate split between himself and the other; in ultimatum the receiver also has a say in that he can either accept or reject the proposal – in the

[1] For a brief overview of different distributive justice principles see Lissowski 2001, pp. 29–38.

ISBN 83-7431-015-4 ISSN 0860-150X

latter case a proposer loses his entire initial endowment while a receiver gets nothing; and a gangster game is an explicit reversal of a dictator – it's the receiver who takes ultimate decision on how much to take from the initial owner of the cake. Dictator and gangster games have also been combined to form a democracy game, where a number of 'haves' and 'have-nots' vote over the final split of the pie. In this kind of game both forced and voluntary redistribution is brought about by a political decision procedure.

Redistribution game with charity transfers

The aim of the article is to propose a redistribution game, well-fit to distinguish between voluntary and politically-enforced tax transfers, as well as to model simple dynamics of such a twofold redistributive process.

The initial distribution of payoffs is exogenous to the game, though it may rightly be thought of as resulting from free-market labor contracts. Except for differences in amount of their initial earned incomes, agents are equal with regard to the type of decision each of them is to make. First, they take part in voting procedure on equal terms. Second, they may dispose their after-tax money in any manner suitable to them by keeping an arbitrary share of their income to themselves, and spreading the rest of it to the others (given they did not keep the entire sum to themselves).

Redistribution through tax system is forced in a sense that once a tax-rate is decided upon, agents are forced to pay a given percentage of their income, irrespectively of their own opinions on the right level of taxation. The presence of coercive element in tax collection renders it necessary to reserve some money for covering the cost of executing taxes from the reluctant. The cost of taxation (C) is defined as a given percentage of total tax revenues, and it may assume any value between 0% and 100% (or more conveniently between 0 and 1). $C = 0$ would imply absence of any executive cost, and $C = 1$ would amount to all-prodigal system in which the whole tax revenues are used exclusively to defray the costs of their collection.

Knowing an exact executive cost C, each player is called upon to cast a personal tax vote (t_i), which likewise may assume any value between 0 and 1. Then a linear tax-rate T is determined by a democratic rule as an average of all players' proposals:

$$T = \frac{1}{n}\sum_{i=1}^{n} t_i$$

The sole dedication of fiscal system in our model is diminishing the existing payoff inequalities. After collecting income taxes proportional to agents' initial earnings and pooling them into the budget, the executive cost C is subtracted, and the remaining sum is equally divided among the players. Two extremes would be $T = 0$, i.e. a *laissez-faire* system in which nobody pays any taxes whatsoever and each player is left with his initial payoff, and $T = 1$, i.e. a *strict egalitarian* system, where all incomes are taken away from the agents and subsequently divided equally between them. Obviously both laissez-faire and strict egalitarian systems can come about only as a result of all players voting 0 or 1, respectively.

If we denote player i's initial payoff as p_i, then his after-tax payoff p'_i, allowing for actual difference between tax paid and subsidy received, can be computed as follows:

$$p'_i = (1 - T)p_i + T(1 - C)\bar{p}$$

where $\bar{p}$ is a mean primary payoff over all players.

Holding tax-rate and cost constant, player i's after-tax payoff p'_i depends partly on his own initial income p_i (the first summand) and partly on a mean primary payoff of a group $\bar{p}$ (the second summand). After-tax incomes for some characteristic combinations of cost and tax level are juxtaposed in Table 1.

Table 1

Player i's after-tax payoff

		Cost of execution		
		0.0	0.5	1.0
Tax-rate	0.0	p_i	p_i	p_i
	0.5	$\frac{1}{2}p_i + \frac{1}{2}\bar{p}$	$\frac{1}{2}p_i + \frac{1}{4}\bar{p}$	$\frac{1}{2}p_i$
	1.0	$\bar{p}$	$\frac{1}{2}\bar{p}$	0

As it was already mentioned, in laissez-faire system ($T = 0$) there is no redistribution and all agents retain their primary payoffs, while in strict egalitarian system ($T = 1$) all players receive the same amount, dependent on the average initial income and cost of tax execution.

Now if we take into account the difference between the initial and after-tax payoff, it can be noted that as a matter of fact the tax system is not linear. All transfers to and from budget included, real lump-sum of a tax paid by player i under tax-rate T is given by the formula:

$$\tau_i(T) = p_i - p_i' = T(p_i - (1 - C)\bar{p})$$

For any T, all agents whose relative initial payoff ($p_i/\bar{p}$) is higher than $1 - C$ pay positive taxes ($\tau_i > 0$), while those whose relative initial payoff is below that threshold receive net subsidy from the budget ($\tau_i < 0$). People initially earning exactly $(1 - C)\bar{p}$ can neither gain nor lose from the tax redistribution[2]. Given this, it is trivial to determine how self-interested players should vote in order to establish linear tax-rate T that would maximize their after-tax payoffs. The whole society of players essentially splits into two groups, which for convenience reasons we shall call 'rich' ($\frac{p_i}{\bar{p}} > 1 - C$) and 'poor' ($\frac{p_i}{\bar{p}} > 1 - C$). It is in direct interest of the rich to vote for zero-percent tax, while in the direct interest of the poor it is to opt for 100% redistribution. The 'intermediate class' has no interest at all in any concrete tax-rate. To avoid random voting, we may assume that all agents having maximized their own after-tax income, in the second place vote to maximize an average after-tax income of a community as a whole. This would lead 'intermediates' to vote in line with a non-redistribution principle self-interested voting scheme, in which we shall make a point of reference for further analysis, is presented in Figure 1.

It is clear that as the cost increases, the threshold value for relative initial income $p_i/\bar{p}$ drops from 1 (for $C = 0$) to 0 (for $C = 1$). Thus for any given initial distribution of income a number of players championing a complete redistribution is a non-decreasing function of C. As the cost is approaching its absolute maximum at 1, the tax-rate T established by a popular vote in a society of self-interested players will be closer to 0. In an extreme case where $C = 1$, virtually no agent can gain from tax redistribution, and a laissez-faire system must prevail. Generally, we will refer to the tax-rate T established in a society of egoistic players as to the *Polarized Voting Tax* (PVT)[3].

As opposed to taxes, redistribution through individual charitable transfers to other players involves no costs as there is no need to compel people to do what they are willing to do of their own initiative[4]. Thus if we

[2] This may be seen as an exemplification of positive/negative income tax, advocated by Milton and Rose Friedman (1996, pp. 114–119).

[3] Similar redistribution mechanism to the one described above, though not allowing for voluntary charity transfers, was incorporated by Elizabeth Jean Wood (1999) into her model of rapid social change. Prof. Wood defined C to be an increasing function of T, and analyzed the model from a point of view of the decisive voter.

[4] To be sure, it is simplifying assumption as private charity also incurs cost of collecting and distributing donations. However, to justify this feature of our model it is enough to notice that, as empirical evidence shows, the cost of private charity is substantially lower than in state-administered system (West, Ferris, 1999).

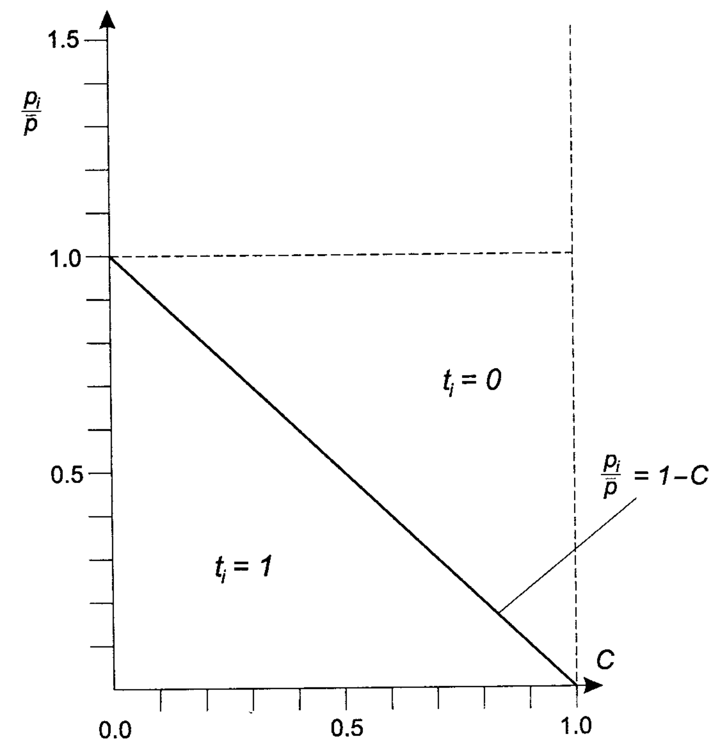


Figure 1. Polarized voting

denote charity transfer player i gives away as $\hat{c}_i$, and charity transfer player j receives as $\check{c}_j$, we can state that:

$$\sum_{i=1}^{n} \hat{c}_i = \sum_{j=1}^{n} \check{c}_j$$

In the following for simplicity reasons we will focus our attention on the redistribution process in a dyadic society with one rich and one poor player. However, it should be noted that this restriction obviously stops us from investigating some distinct new qualities of the game that emerge as number of agents exceeds 2.

Game with 2 players

Let us start with presenting a one-shot normal form redistribution game for two players. Suppose that players have different initial payoffs, and cost of execution satisfies following condition:

$$C < \frac{p_R - p_P}{p_R + p_P}$$

where p_P and p_R stand for poor and rich player's payoffs respectively.

The condition warrants that it is in the self-interest of the poor player to vote for 100% tax-rate which will result in establishing *Polarized Voting Tax* (PVT) equal to 50%. If cost of the execution exceeded the threshold value, even a poor player would incur losses at any positive tax level and therefore would be inclined to vote 0%.

It is obvious that at non-zero cost of the tax execution rich agent cannot opt for anything but 0% tax. Even if he is an altruist willing to support a poor player, it is better for him to do it by direct charitable payment which does not involve any cost. It is also plausible to assume that no poor agent will be interested in passing part of his initial payoff to the rich as this would further augment the original gap between himself and the latter. Therefore, charity transfer poor player gives away to the other in no case exceeds zero. Under these assumptions, rich votes 0% and chooses between alternative amounts of charitable transfer, while the poor one gives no donation and chooses how to vote between alternative tax-rates. Resulting from global tax-rate, as it is half the tax-rate proposed by the poor, varies between 0 and 50% (PVT).

With the help of the example suppose that initially a rich player earns $25 and poor $15, while the cost of execution is 10%. Suppose further that the rich considers giving $2 to support the worse-off player, and the latter takes into account voting either 0 or 100%. Such a game can be presented in the following table:

Table 2

Redistribution game in normal form

		Poor	
		$t_P = 0 \to T = 0.0$	$t_P = 1 \to T = 0.5$
Rich	$\hat{c}_R = 2$	17.00 23.00	18.50 19.50
	$\hat{c}_R = 0$	15.00 25.00	16.50 21.50

Both people have dominant strategies: regardless of poor player's choice it is better for the rich player not to make any donation, and regardless of rich player's behavior it is more profitable for the poor to establish maximum 50% tax-rate by voting 100%. Thus, the equilibrium solution is for the rich

player not to contribute anything, and for the poor to vote for maximum redistribution possible (in which case the rich person's final payoff is $21.50, whereas for the poor it is $16.50). This, however, is not a Pareto-optimal result as for both players, it would be more profitable, had the rich chosen to donate $2, and the poor decided to opt for a laissez-faire system (then the rich would have $23 and the poor $17). Thus, a redistribution game turns out to be an asymmetric variant of prisoner's dilemma.

The surplus players can divide between themselves the results from reduction of execution cost due to the virtual elimination of tax system. Holding players' initial payoffs constant, the lump-sum of the surplus depends on the actual tax-rate and the cost of execution.

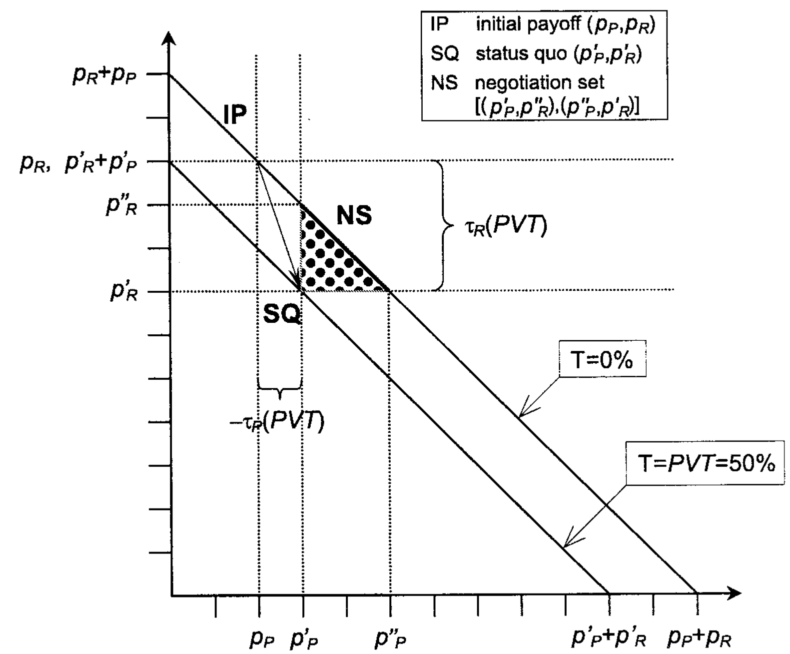


Figure 2. Negotiating surplus

In Figure 2 we see feasible outcomes of the game if players are in position to conclude a binding contract (poor and rich agent's payoffs at horizontal and vertical axes, respectively). They start the game with their initial payoffs (*IP*) when the rich earns p_R and the poor p_P. Under a democracy rule, a poor player can redistribute some wealth from the rich to himself by raising tax-rate up to 50% (PVT). By-product of tax redistribution is shifting of the budget constraint down and to the left. The shift occurs because at $T = 0.5$ some part of players' total payoff is consumed by the executive cost. Now we may consider a point (p'_P, p'_R) to be status quo (SQ) since it is a pair of incomes that each agent is able to assure himself of his own, regardless of the other's behavior. At SQ the rich pays lump-sum tax equal to $\tau_R(PVT)$, and the poor receives a net subsidy equal to $-\tau_P(PVT)$. The difference between the two is the amount that could be re-gained by establishing laissez-faire system. However, the poor has no direct interest in lowering tax-rate for it would shift the outcome back in direction of IP, thus reducing his final payoff. On the other hand, 50% tax-rate brings harm to the rich player who not only covers the subsidy to the poor but also defrays the entire executive cost. Thus, it would be much to his interest to replace a politically-forced costly redistribution with a voluntary cost-free charity transfer to the other. To encourage the poor to vote for 0% tax-rate, though, the rich player needs to offer him a lump-sum at least equal to the loss incurred by the poor from eliminating tax redistribution, i.e. $-\tau_P(PVT)$. The thick black line in Figure 2 indicates the negotiation set, i.e. a number of solutions to the problem of how the surplus gained from abolition of tax system should be divided between the players. If entire surplus goes to the rich, the ultimate outcome will be (p'_P, p''_R), if it falls solely to the poor, the outcome will be (p''_P, p'_R). All combinations of payoffs between those two points are feasible as well.

Dynamical substitution between voluntary and forced redistribution

At this point we shall introduce dynamics into the system. Suppose the game is infinitely iterated, with initial payoffs and the cost of tax execution held constant and known to the players who have no possibility of direct communication. Each round will consist of the following sequence of moves: first players vote on redistribution, then tax transfers to and from the budget take place according to the current tax-rate, and finally it is up to players to give away some part of their income to the other.

At the outset poor player, willing to secure to himself the status quo outcome, votes for maximum redistribution and thus PVT at 50% is established. Now the rich player has an occasion to signal his willingness to replace tax redistribution with voluntary transfer by offering a donation to the poor. The lump-sum of this first donation depends on rich agent's *charitable initiative*. In the second round the poor player may react to the donation by reducing his demand for tax redistribution according to his personal *demanding attitude*. In turn the rich player reacts to

tax-rate decrease according to his *generosity*[5], and so forth the process continues ad infinitum. Let us now formally define the individual features of players:

- ε – *charitable initiative*, is an amount of income, expressed as a share of $\tau_R(PVT)$, that a rich player is willing to give away directly to the other in the first round of the game; ε is effective as long as it satisfies the condition $\varepsilon \le \frac{p_R - \tau_R(PVT)}{\tau_R(PVT)}$; greater values of ε are cut down to that threshold level, for donation cannot exceed rich agent's entire income.
- γ – *generosity*, is a share of $\tau_R(PVT)$ that a rich player is willing to donate under laissez-faire system; effective γ's obey $\gamma \le \frac{p_R}{\tau_R(PVT)}$.
- δ – *demanding attitude*, is a share of $-\tau_P(PVT)$ that a poor player will demand as a compensation for eliminating tax redistribution completely[6].

Algorithm of decision making in each round of the game may be summarized in the table 3.

As we see in each round the poor player gives no donation ($\hat{c}_P = 0$), and the rich player votes for 0% tax-rate ($t_R = 0$). From the second round on tax-vote by the poor agent depends on his personal demanding attitude γ and the donation received in the preceding round. If at time $i-1$ he received no charity transfer, at time i he will vote 100% (except for $\delta = 0$, in which case he votes 0% regardless of donation received); if transfer equaled $-\delta\tau_P(PVT)$ or more, he will be inclined to vote 0% (t_P's lower limit is obviously 0, so lower values are automatically increased to this level); if donation was somewhere between 0 and $-\delta\tau_P(PVT)$, he will depart from 100% vote proportionately. It is safety strategy for the poor to have $\delta \ge 1$. In case $\delta \ge 1$, he is ready to vote 0% only if rich agent fully covers the loss incurred by the poor from abandoning the status quo.

On the other hand, rich agent's donation depends on his individual generosity γ and tax-rate T in the current round. If tax-rate is PVT, he

[5] These three features (parameters) define a type of agent, though they are activated contextually, i.e. for a player in poor position only demanding attitude is relevant, whereas for a player in rich position it is charitable initiative and generosity. It should also be stressed that we use the terms in neutral sense and attach no moral value, neither positive nor negative, to any of the features. For instance, 'generous' actions may as well be motivated by strict self-interest.

[6] Negative sign before $\tau_P(PVT)$ is due to the fact that at PVT poor player by definition 'pays' negative tax (which means that actually he receives net subsidy from the budget).

Table 3

Dynamics of redistribution game[7]

		Poor player	Rich player
Round no. 1	Voting	$t_P = 1$	$t_R = 0$
	Donation	$\hat{c}_P = 0$	$\hat{c}_R = \varepsilon\tau_R(PVT)$
Round no. 2 and next	Voting	$t_P = \begin{cases} 0 & \delta=0 \\ 1 + \frac{\hat{c}_P}{\delta\tau_P(PVT)} & \delta>0 \end{cases}$ * $\hat{c}_P = \hat{c}_R$ of the preceding round	$t_R = 0$
	Donation	$\hat{c}_P = 0$	$\hat{c}_R = \begin{cases} 0 & PVT=0 \\ \frac{PVT-T}{PVT} \cdot \gamma\tau_R(PVT) & PVT>0 \end{cases}$ * T of the current round

gives no donation, regardless of his generosity; if T is 0%, his donation equals $\gamma\tau_R(PVT)$; if T is somewhere between 0 and PVT, he offers a proportionate charity transfer. His safety strategy is to have $\gamma \leq 1$. In case $\gamma = 1$, he is ready to donate to the poor entire surplus resulting from abolishing tax redistribution. Was his generosity greater than that, he would expose himself to the risk of earning less than status quo.

Under what conditions can tax redistribution be completely replaced with voluntary transfers on part of the rich? To answer this, let us find levels of ε, γ, and δ that make it possible to establish laissez-faire system. Roughly speaking, to make a tax-rate go down to zero, it is necessary that the rich was generous *enough*, whereas the poor was not demanding *too much*. A threshold value of poor's demanding attitude (δ) as a function of rich's generosity (γ) is given by the *correspondence formula*[8] (see Appendix for details):

$$\delta = -\frac{\tau_R}{\tau_P} \cdot \gamma$$

[7] The same algorithm may be applied to multiplayer game. Tax-vote by poor player depends on the donation he received in the previous round, though when number of players exceeds two, the donation need not be equal to the transfer made by any particular rich agent (one possible way to distribute charity transfers among the poor is to employ leximin principle). The 'intermediate class' could be defined to have $t_I = 0$ and $\hat{c}_I = 0$ for each round.

[8] The visual presentation of the equation will be referred to as *correspondence line*. To avoid excessive notation, from now on τ_R and τ_P will denote lump-sum tax paid at PVT rate.

If δ satisfies the equation than substitution rate between taxes and free donations is exactly the same for both agents. It means that in order to completely eliminate taxes the rich player is ready to give away the amount that is precisely as much as poor player demands for reducing his tax-vote to zero. If δ exceeds the threshold value, tax-rate T will be equal to PVT from the very beginning, or will be approaching limit at PVT with speed negatively correlated to the charitable initiative of the rich (ε). If δ is less than that, at some point of the game laissez-faire system will be established (the smaller δ and the larger ε, the sooner tax-rate will fall to zero). If δ is exactly equal to the threshold value, than the tax-rate established in the second round will hold for the rest of the game. Thus ultimate tax-rate in this case is determined by ε and may assume any value between 0 and 1. Its exact value is given by (see Appendix):

$$T = \tfrac{1}{2}\left(1 - \frac{\varepsilon}{\gamma}\right), \quad \gamma > 0,\ \varepsilon \leq \gamma$$

Payoff structure at time approaching infinity

To conclude the analysis we will examine limit payoffs, as time approaches infinity, for different combinations of generosity γ and demanding attitude δ, while holding charitable initiative ε equal to γ[9]. Next we will point pairs of γ and δ that form Nash equilibria. To make the analysis easier to follow we will plot the results on the same diagram as seen in Figure 3.

Let us denote the rich agent's status quo payoff (PVT, no donation) as R0, payoff better than that as R+, and worse one as R–; and respectively for poor agent: P0, P+, P–. In Figure 4 we show how given combinations of γ and δ prove better/worse than status quo for either player.

If generosity γ of the rich agent is too small in comparison to demanding attitude δ of the poor, the tax-rate will approach a limit at PVT, and thus both players will receive their status quo payoffs R0, P0 (entire area above correspondence line). However, if γ and δ are kept in 'reasonable' proportion to each other, a lower tax-rate is established, and there is a surplus resulting

[9] For a rich agent playing his safety strategy ($\gamma \leq 1$), it is always profitable to have laissez-faire system established. As we focus on the limit payoffs at infinity, single first round charity expense determined by ε does not affect rich agent's payoff, so he does not incur any more risk by giving ε value equal to (or even higher than) δ. And by doing so he is able to reduce tax-rate to zero in case γ and δ happen to lie on the correspondence line.

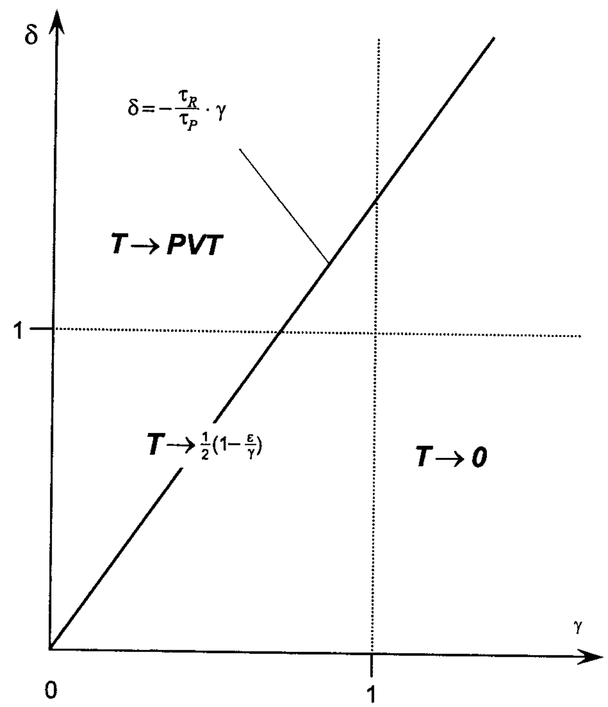


Figure 3. Generosity, demanding attitude and tax-rate

from cutting down on execution costs. Thus on and below the correspondence line always at least one of the agents is better-off than in status quo. It's worth noticing that for any given value of γ both agents' limit payoffs are independent of δ, as far as δ does not exceed the threshold value[10]. Particularly, for $\gamma = -\tau_P/\tau_R$ entire surplus is taken by the rich, while the poor is left with his status quo earning, and for $\gamma = 1$ entire surplus goes to the poor, while the rich keeps his status quo income. Generally, for all points on and below the correspondence line, as γ coordinate is reduced, the rich agent's payoff increases 'at the expense' of the poor.

We may now see that when considering limit payoffs redistribution game is basically a variant of ultimatum game. The rich agent proposes a donation depending on his generosity γ, and the poor agent either accepts the offer or rejects it depending on his demanding attitude δ. In case δ lies above a correspondence line and the proposition is rejected, both players are left with their status quo payoffs. In contrast to the original ultimatum game

[10] The reason for it is that parameter γ by its very definition determines the lump-sum that goes to the poor by means of voluntary transfer under laissez-faire system, and as we know from Figure 3 the limit tax-rate under correspondence line is zero.

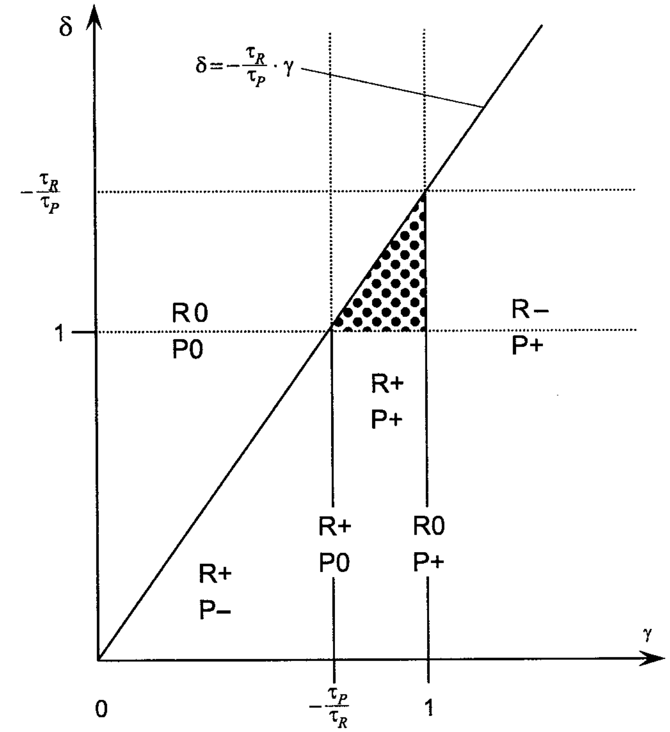


Figure 4. Limit payoffs in comparison to status quo

where players' payoffs could not fall below zero, reaching an 'agreement' in the redistribution game may lead to one of the players towards being materially worse-off than in status quo. The latter situation is possible when either player does not play his safety strategy.

As we may read from the diagram, safety strategy for the poor player is to have $\delta \geq 1$. Was δ below 1, the poor could suffer loss in comparison to status quo, if the rich had his γ below $-\tau_P/\tau_R$. Similarly it is safe for the rich to have $\gamma \leq 1$. Was his generosity greater than that, his limit income could fall short of status quo, in case the poor was not too demanding ($\delta \leq -\tau_R/\tau_P$). The dotted triangle indicates pairs of safe γ and δ that bring profit to both sides (except for point $(-\tau_P/\tau_R, 1)$ and the right side of the triangle, where only one player gains, while the other stays with status quo payoff). All γ–δ pairs in the triangle (and rectangle below it as well) lead to payoffs that belong to the negotiation set presented in Figure 2.

Further scrutiny leads us to the conclusion that for a self-interest maximizing poor player $\delta = 1$ is a dominant strategy. Such a choice is the analogue of the receiver accepting zero-share in ultimatum game with continuous payoffs.

Nash equilibria of the redistribution game

Finally, let us point out combinations of γ and δ that form Nash equilibria of the redistribution game. First of all, we may rule out all points that violate either agent's safety level, i.e. $\gamma > 1$ or $\delta < 1$. Further let us consider γ belonging to $(-\tau_P/\tau_R, 1]$. The poor player's best response to any given γ in the range is **any** δ less or equal to threshold value. However, rich agent's best response for any δ less or equal to $-\tau_R/\tau_P$ is choosing γ in such a way as to locate the point (γ, δ) on the correspondence line. Thus all points on a hypotenuse of a dotted triangle in Figure 4 are Nash equilibria (thick line in Figure 5), while the others are not[11].

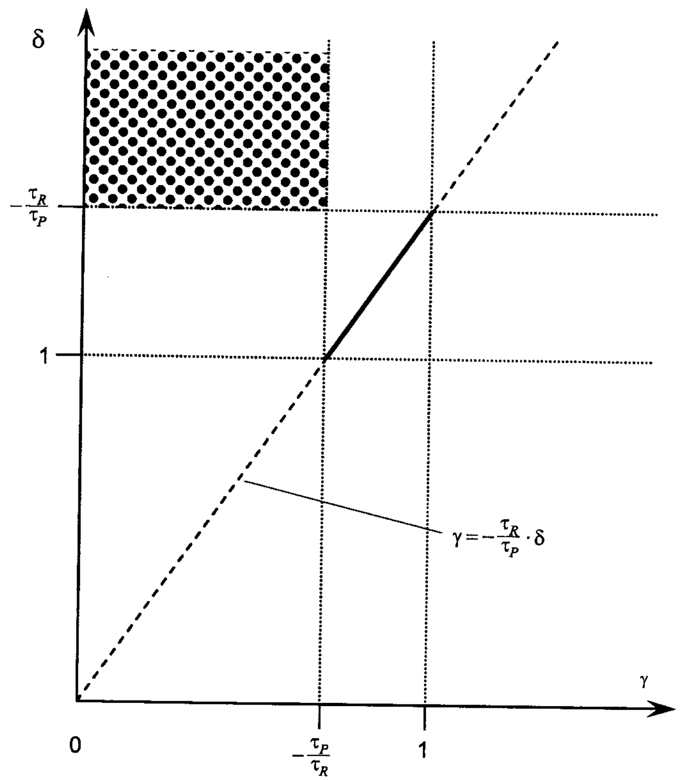


Figure 5. Nash equilibria (for $\varepsilon = \delta$)

[11] As $\delta = 1$ is a poor player's dominant strategy, combination of $\delta = 1$ and $\gamma = \varepsilon = -\tau_P/\tau_R$ is a unique solution of the redistribution game for strictly self-interested players. In this case poor player gets exactly his status quo income while the rich increases his status quo payoff by re-gaining the entire surplus resulting from abolition of executive costs of tax system.

For γ belonging to $[0, -\tau_P/\tau_R]$ any level of δ over the correspondence line may be considered the best response, as well as for δ equal to or greater than $-\tau_R/\tau_P$ the best response could be any γ to the left of the correspondence line. Thus Nash equilibria are also all points of the dotted rectangle, although they obviously are not Pareto-optimal. As empirical evidence from dictator and gangster games shows, it may well be the case that subjects actions lead to suboptimal outcomes. "The generous nature of individuals found in fairness games does not overcome the distribution struggle. While dictators are prepared to give up a part of their endowment, gangsters demand a much bigger share of the cake for themselves" (Eichenberger Oberholzer-Gee 1998, p. 196).

Graphical illustration of system dynamics

To give an example of redistributive dynamics we will conclude the article with a few characteristic cases of system evolution. In each case agents' initial payoffs are \$90 (rich) and \$10 (poor), whereas the cost of tax execution is 20%. At *PVT* (50%) the rich agent's payoff is equal to \$65 and the poor player earns \$25. On each graph a line - - - refers to the rich agent's final income (all taxes and voluntary transfers included), a line ---- to the poor agent's income, and a line -.-.- reports the history of the tax-rate.

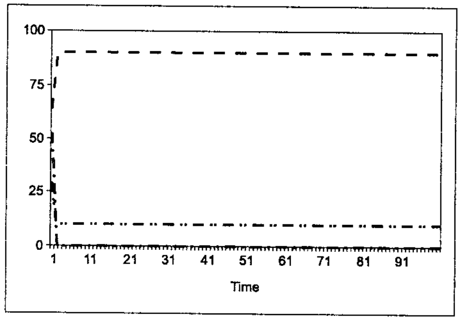
Illustration no. 1 shows how a stable state is immediately reached if the poor agent exhibits no demand for redistribution ($\delta = 0$). If at the same time the rich agent exhibits no generosity ($\gamma = 0$), he takes entire surplus resulting from eliminating executive costs, and both players earn their initial payoffs. It may be seen as an instance of the poor player's high moral standards that hold him back from exploiting the democratic procedure to enforce more profitable income distribution. To be sure, that sort of consideration would not even pass through the mind of *homo œconomicus*.

Illustration no. 2 shows that if the poor agent's demanding attitude is anything apart from zero, it is impossible to establish laissez-faire system without charitable initiative of the rich ($\varepsilon = 0$) – no matter how great his generosity could be ($\gamma \geq 0$). The tax-rate does not deviate from *PVT* even by the smallest margin and in each round the agents earn their status quo payoffs (\$65 and \$25).

Graphs nos. 3 and 4 show the history of reaching Pareto-optimal distribution (with both agents' payoffs higher than status quo). In no. 3 the rich player does not play his best response. By reducing his generosity and

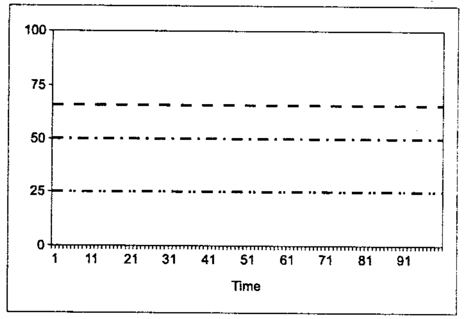
1. No demand for redistribution

γ=0.00 ε=0.00 δ=0.00



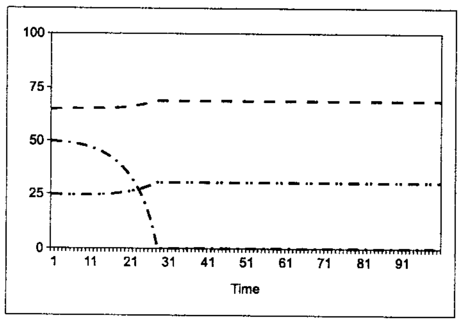
2. No charitable initiative (suboptimal NE)

$\gamma \geq$0.00 ε=0.00 δ>0.00



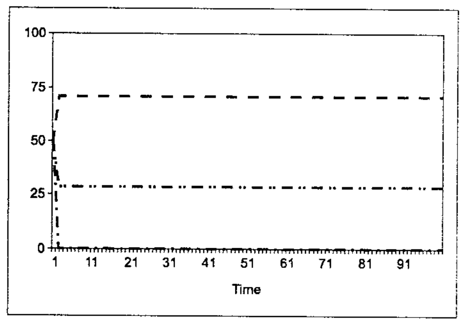
3. Reaching Pareto-optimal outcome (not NE)

γ=0.85 ε=0.01 δ=1.25



4. Pareto-optimal Nash Equilibrium

γ=0.75 ε=0.75 δ=1.25

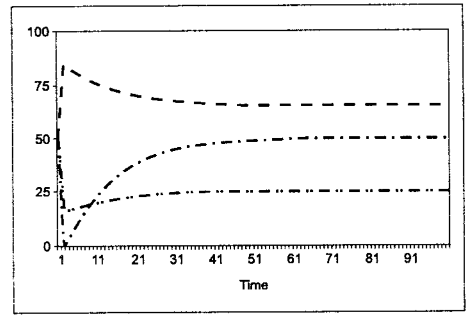


raising his charitable initiative appropriately, he could assure himself a higher payoff, taking benefit of the poor agent's moderate demand for redistribution. If he did so, Nash equilibrium presented in graph no. 4 would be established: (γ = 0.75, δ = 1.25) is a point lying exactly on the correspondence line.

Graphs nos. 5 and 6 show how system can recede to *PVT* after initial reduction of tax-rate to nigh-zero level. No. 5 illustrates that it is impossible to dupe the poor player into laissez-faire system by substantial charitable initiative combined with low generosity. Even though the poor agent plays δ far below his safety strategy, the tax-rate after the initial reduction gradually recedes to *PVT*. On the other hand, in no. 6 poor player's demand for redistribution is too much even for a super-generous rich agent. In spite of large initial charity transfers (that even made the recipient wealthier

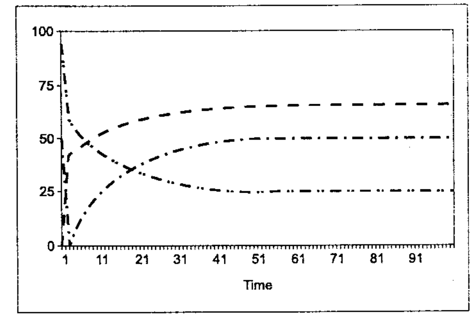
5. Recession to PVT due to insufficient generosity

γ=0.20 ε=0.40 δ=0.36



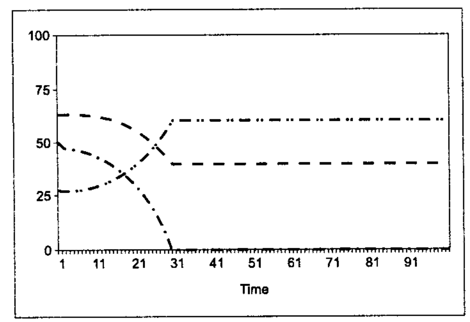
6. Recession to PVT due to excessive demand

γ=2.00 ε=2.50 δ=3.60



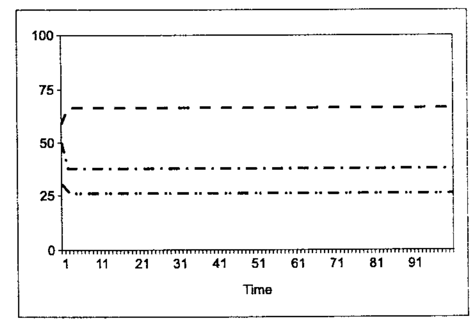
7. Super-generosity

γ=2.00 ε=0.10 δ=3.00



8. Insufficient charitable initiative

γ=0.80 ε=0.20 δ=1.33



than his benefactor), the poor player departs from voting 0% and *PVT* is gradually brought back.

Illustration no. 7 depicts a rich man who is willing to give away most of his income, starting with a small initial contribution. The poor agent's demanding attitude is below the threshold value, so the tax-rate is being gradually reduced to zero. The poor agent takes benefit of the rich agent's generosity, receiving a substantial part of his wealth, and ultimately becoming richer than the donor.

Finally, at no. 8 we see a game between two players whose generosity and demanding attitude are exactly corresponding to each other. However, due to insufficient charitable initiative on the part of the rich, the tax-rate is fixed at 37.5% and benefits from complete eliminating executive costs are lost.

Concluding remarks

The purpose of the article was to model a redistribution process, allowing for the interplay between transfers forced by means of the tax system and voluntary donations to the worse-off. The dynamics of the system were guided by agents' personal features, namely charitable initiative and generosity on part of the rich, and demanding attitude on part of the poor. We have shown that even under the assumption of exclusive self-interest seeking, there are Pareto-optimal Nash equilibria that result in a complete substitution of free charity for tax redistribution, as well as suboptimal equilibria that keep the volume of tax redistribution intact. It is a question of empirical research whether real-life subjects are able to find their way to elimination of excessive cost of politically forced transfers. It is also a matter for further discussion how dynamics of the game are affected by introducing greater number of players and focusing attention on discounted payoffs rather than looking at the limit distribution of income.

Appendix

To simplify notation let the variables over time be denoted as:

c_i – voluntary transfer from rich to poor player in round i,
t_i – poor agent's tax-vote in round i,
T_i – tax-rate in round i,

Constants (as described in the text):

$p_R, p_P, \bar{p}, \tau_R, \tau_P, C, \varepsilon, \gamma, \delta$

At time i player k's payoff is given by formula:

(1) $$p_k(T_i) = (1 - T_i)p_k + T_i(1 - C)\bar{p}$$

According to decision algorithm (see Table 3), rich agent's donation in first round is given by:

(2) $$c_1 = \varepsilon\tau_R$$

His donations in round $i \geq 2$ are:

$$c_i = \frac{PVT - T_i}{PVT}\gamma\tau_R.$$

Since $PVT = \frac{1}{2}$, we get

(3) $$c_i = (1 - 2T_i)\gamma\tau_R.$$

Tax-rate at time i is always half the tax-rate proposed by the poor (again see Table 3):

(4) $$T_i = \tfrac{1}{2}t_t = \tfrac{1}{2}\left(1 + \frac{c_{i-1}}{\delta\tau_P}\right)$$

Substituting (2) into (4), we obtain tax-rate in second round:

$$T_2 = \tfrac{1}{2}\left(1 + \frac{\varepsilon\tau_R}{\delta\tau_P}\right).$$

Substituting (3) into (4), we obtain difference equation for $i \geq 3$:

$$T_i = -\frac{\gamma\tau_R}{\delta\tau_P}T_{i-1} + \tfrac{1}{2}\left(1 + \frac{\gamma\tau_R}{\delta\tau_P}\right).$$

To make it easier to handle let us rewrite it as:

(6) $$T_i = aT_{i-1} + b,$$

where $a = -(\gamma\tau_R/\delta\tau_P)$, $b = \frac{1}{2}(1 - a)$.

Or, alternatively:

(7) $$T_i = a(T_{i-1} - \tfrac{1}{2}) + \tfrac{1}{2}.$$

Now the sequence of T_i's for $i \geq 3$ is:

$$T_3 = aT_2 + b$$
$$T_4 = aT_3 + b = a^2T_2 + ab + b$$
$$\vdots$$
$$T_i = a^{i-2}T_2 + a^{i-3}b + a^{i-4}b + \ldots + ab + b = a^{i-2}T_2 + \tfrac{1-a^{i-2}}{1-a}b =$$
$$= a^{i-2}T_2 + \tfrac{1}{2}\left(1 - a^{i-2}\right) = a^{i-2}\left(T_2 - \tfrac{1}{2}\right) + \tfrac{1}{2}$$

Thus for $i \geq 3$:

(8) $$T_i = a^{i-2}(T_2 - \tfrac{1}{2}) + \tfrac{1}{2}.$$

Solving (8) with $i \to \infty$ we get:

1. $0 \leq a < 1$:
   $a^{i-2} \to 0$ and $\lim_{i\to\infty} T_i = \frac{1}{2}$
2. $a = 1$:
   $T_i = T_2 \quad \forall i$
3. $a > 1$:
   if $T_2 = \frac{1}{2}$ (which implies $\varepsilon = 0$ and $\delta > 0$): $T_i = \frac{1}{2} \quad \forall i$,
   if $0 \leq T_2 < \frac{1}{2}$: $\exists m \geq 2$ for which $T_m \leq 0$ (since in (8) the expression in parentheses is negative, and a^{m-2} increases in m). Since any T lower than zero is automatically increased to zero, we have $T_m = 0$.

Now let m be the smallest possible number of a round. In accordance with (7), if $T_m = 0$, then $T_{m+1} = (1-a)/2 < 0$. Therefore $T_{m+1} = 0$, and since the same holds true for all subsequent rounds, $\lim_{i\to\infty} T_i = 0$.

Deciphering a, we get:

1. For $\delta > -\frac{\tau_R}{\tau_P} \cdot \gamma$, $\lim_{i\to\infty} T_i = \frac{1}{2}$, $\gamma > 0$. If $\gamma = 0$, $T_i = 0\ \forall i \geq 2$ (see Table 3).
2. For $\delta = -\frac{\tau_R}{\tau_P} \cdot \gamma$, $\lim_{i\to\infty} T_i = T_2 = \frac{1}{2}\left(1 + \frac{\varepsilon\tau_R}{\delta\tau_P}\right)$, $\delta > 0$. Since $\delta = -\frac{\tau_R}{\tau_P} \cdot \gamma$, after transformation we receive: $\lim_{i\to\infty} T_i = \frac{1}{2}\left(1 - \frac{\varepsilon}{\gamma}\right)$, $\gamma > 0$. If $\delta = \gamma = 0$, $T_i = 0\ \forall i \geq 2$ (see Table 3).
3. For $\delta < -\frac{\tau_R}{\tau_P} \cdot \gamma$,
   a) if $\varepsilon = 0$ and $\delta > 0$, $\lim_{i\to\infty} T_i = \frac{1}{2}$;
   b) if $\varepsilon > 0$, $\lim_{i\to\infty} T_i = 0$.

References

Bolton, Gary E., Axel Ockenfels, "Self-centered fairness in games with more than two players", forthcoming in *Handbook of Experimental Economics Results*, http://www.mpiew-jena.mpg.de/esi/discussionpapers/paperspdf/2002-42.pdf.

Eichenberger, Reiner, Felix Oberholzer-Gee, 1998, "Rational Moralists: The Role of Fairness in Democratic Economic Politics", Public Choice 94, pp. 191–210, http://www.unige.ch/ses/ecopo/staff/robert/Eichenberger_OberholzerGee_1998_pc.pdf.

Fehr, Ernst, Klaus Schmidt, 1999, "A Theory of Fairness, Competition and Cooperation", *The Quarterly Journal of Economics*, August 1999, pp. 817–868, http://e-collection.ethbib.ethz.ch/ecol-pool/incoll/incoll_518.pdf.

Fong, Christina M., 2003, "Empathic Responsiveness: Evidence from a Randomized Experiment on Giving to Welfare Recipients", http://bpp.wharton.upenn.edu/Acrobat/Fong_AEW_Paper_4_16_03.pdf.

Friedman, Milton, Rose Friedman, 1996, "Wolny wybór", Wydawnictwo PANTA, Sosnowiec.

Gächter, Simon, Arno Riedl, 2002, "Moral Property Rights in Bargaining", Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 113, http://e-collection.ethbib.ethz.ch/ecol-pool/incoll/incoll_624.pdf.

Levine, David K., 1997, "Modeling Altruism and Spitefulness in Experiments", http://levine.sscnet.ucla.edu/Papers/bilat.pdf.

Lissowski, Grzegorz, 2001, "Problemy i metody teorii wyboru społecznego", w: Grzegorz Lissowski (red.), *Elementy teorii wyboru społecznego*, pp. 9–46, Wydawnictwo Naukowe SCHOLAR, Warszawa.

Neugebauer Tibor, Anders Poulsen, Arthur Schram, 2002, "Fairness and Reciprocity in the Hawk-Dove Game", Aarhus School of Business, Department of Economics, Working Paper 02–12, http://www.hha.dk/nat/WPER/02-12_aup.pdf.

Rutström, E. Elisabet, Melonie B. Williams, 2000, "Entitlements and Fairness: An Experimental Study of Distributive Preferences", http://dms-web.moore.sc.edu/lisa/research/equity_feb00.PDF.

Tyran, Jean-Robert, Rupert Sausgruber, 2002, "A Little Fairness May Induce a Lot of Redistribution in Democracy", University of St. Gallen, Department of Economics, Discussion Paper no. 2002–30, http://www.vwa.unisg.ch/RePEc/usg/dp2002/dp0230Tyran_Sausgruber_ganz.pdf.

West, Edwin G., J. Stephen Ferris, 1999, "Private versus Collectivized Charity: Further Explorations of the Crowding Out Debate", http://www.carleton.ca/economics/cep/cep99-01.pdf.

Wood, Elisabeth Jean, 1999, "The Stakes of the Game: The Politics of Redistribution in Democratic Transitions", http://www.yale.edu/leitner/pdf/wood.pdf.

Simon Czarnik
Jagiellonian University
scisuj@o2.pl

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 7 (20) 2004

Mikołaj Jasiński
Warsaw University

NICE, CONVENTION, COMPROMISES – ABOUT PROCEDURES' MEANING IN PROJECTED POLITICAL INSTITUTIONS

Abstract. This paper discusses problems arising during construction of decision-making rules, using the EU Council's voting system as the main example. The work presents postulates required during a political institution's formal and politically-related construction process. Numerous compromise propositions have appeared during the debate on the voting system for the EU Council including the two main concepts, the Nice and Convention Systems. Here, selected propositions are presented and, using power indices, their consequences are considered. A new concept for attaining compromise is also offered, which requires negotiators to merely establish the power apportionment in the EU, not construct the minutiae of a voting system. The paper presents a way to determine the optimal voting system that satisfies negotiated results as well as postulates presented herein.

I once thought the construction of political decision-making institutions was a matter for politicians or lawyers skilled at formulating appropriately precise principles to regulate decision-making by the entitled. Today, I teach my students about decision-making processes in institutions such as general elections, parliaments and shareholder meetings. We all know that the description of these mechanisms or formulation of normative expectations requires reference to not always intuitive formal theory. Politicians and legal experts negotiating the future decision-making system of the EU Council have completely ignored conclusions of countless expert opinions regarding formal social choice theory; the results of those negotiations evidence exactly the same thing. When considering these results, I formulated a proposed way to construct the voting system in the EU Council that satisfies political as well as formal postulates.

First, I would like to recount the decisional practice of the EU Council, certain traits of which I will attempt to "translate" into formal characteristics. I will also present the expectations of the tool I used to analyze and design the EU Council voting system, which led me to apply the

ISBN 83-7431-015-4 ISSN 0860-150X

Shapley-Shubik index[1]. After describing certain properties of the negotiated EU Council voting system, I will present a method for seeking a compromise voting system. The text refers to the well-known power index properties, described extensively and proven in literature on the subject[2]. However, the ways in which the interpretations of these properties are synthesized to formulate a voting system may be new.

When constructing new institutions, information regarding their goals and the expectations weighing thereon are of prime import. This is rather simple with respect to the EU Council; we are familiar with its decision-making practice and the main principles that define hitherto relations among EU member states. Mentioned practice and prior relations are expressions of the underlying goals of states building the EU. Any transformation of the EU Council's voting system after Union enlargement should thus refer to prior experience and account for these goals. This will allow for the goals and expectations to be expressed in the system. Therefore, I will briefly illustrate the legal framework and historical decision-making practice of the EU Council.

Decision-making in the European Union Council[3]

The EU Council makes its decisions during approximately 80 sessions held each year since the 1990s. The Council works in various configurations, depending on the subjects being considered. Decisions are made by a simple majority, qualified majority or unanimously in meetings of the Union Council. The **qualified majority** system is used most frequently. Simple majority votes are used for the least weighty decisions (organizational and procedural matters), while unanimity of Council members is required in issues of utmost importance such as enlargement of EU membership, amendment of treaties, etc., which occur extremely rarely (it was used more frequently at the outset of the European community's existence).

During qualified majority voting, each state is accorded a certain weight, a number of votes based on its economic, territorial and demographic potential, as well as political issues. The table below presents the allocation

[1] [23], [25].

[2] I have chosen to avoid an excess of formal terminology, limiting myself to the necessary minimum.

[3] I have limited the description of the UE Council's functioning to the necessary minimum. A more thorough description of this subject may be found in nearly every widely available work on the EU.

Table 1

Political weight of EU members at various stages of European integration

Member state	Number of votes				
	6 states	9 states	10 states	12 states	15 states
France	4	10	10	10	10
Germany	4	10	10	10	10
Italy	4	10	10	10	10
Great Britain		10	10	10	10
Spain				8	8
Belgium	2	5	5	5	5
Holland	2	5	5	5	5
Greece			5	5	5
Portugal				5	5
Sweden					4
Austria					4
Denmark		3	3	3	3
Finland					3
Ireland		3	3	3	3
Luxembourg	1	2	2	2	2
TOTAL	17	58	63	76	87
Required qualified majority	12 (70.6%)	41 (70.7%)	45 (71.4%)	54 (71.1%)	62 (71.3%)

of votes at various stages of Union development, beginning with six states (Treaty of Rome, 1957) to the fifteen (fourth enlargement in 1995).

From the very beginning, the vote allocation system assumed a need to reinforce the weaker states' position while seeking to implement solutions that did not deprive larger states of influence. In short, it is a formal expression of interstate solidarity within the Union. Clearly, this is not an easy task. Aside from the issue of whether to maintain the unanimity requirement in the EU Council, establishing the weighting system is one of the most controversial subjects discussed in the united Europe. In 2000, the Nice Treaty (2000) introduced new principles for the Union's Council decision-making. Between November 2004 and October 2009, three conditions will have to be met for a decision to be adopted: 1) the proposal must be supported by more than half the states (this number will be 14 after EU enlargement reaches 27 states), 2) the sum total of weighted state votes must exceed the established threshold (after enlargement, this will be 258 votes, i.e. nearly 75% of all, 345 votes in the entire EU), 3) the total number of residents of states supporting the proposal must constitute at least 62% of the European Union's population. Thus, this is a triple majority system. Note the definite innovation of this voting system. The population factor is added to the traditional "political" weights, which obviously gives large states additional power. This change undoubtedly violates the above-mentioned solidarity expressed in arbitrarily established weights. The second important consequence of the population factor being introduced into the Nice Treaty is the unchallenged violation of the understanding between Gen. Ch. de Gaulle and Chancellor K. Adenauer (prior to the Treaty of Rome on EEC) regarding the invariable equality between France and Germany. The understanding had been treated as one of united Europe's foundations. The new regulation gives the advantage over France to Germany as the most populous state of the European Union.

What is the decision-making practice in the Union's Council within the above formal framework? Above all, the extreme number and multi-dimensionality of this European super-government's decisions stands out: it covers cultural issues, social and health matters, energy policy, farming as well as the Union's budget. The normally daylong sessions have sometimes been extended in crisis or deadlock situations. These "marathon" sessions last until compromises are hammered out. The EU Council attempts to maintain the consensual nature of the decisions adopted in the course of its workings. *Ad hoc* coalitions are usually formed during negotiations; the Union's Council has no permanent coalitions while divisions and interest groups cross all lines.

Because the *ad hoc* coalitions are usually formed during negotiations, it can be assumed that over the long-term (say, over a few years) the ways of conducting negotiations will become extremely diverse and provisional arrangements will be built at random. These observations direct us toward power indices based on an assumption of a large number of decisions. The second observation turns our attention to the extreme diversity of the subject matter, i.e. the multi-dimensionality of the decisions adopted. This diversity allows us to refer to the spatial models of decision-making bodies.

That is all for the legal and procedural aspects of the decision-making process in the Council of the European Union. Further, the specifics of members need to be taken into account in the analysis of the discussed

body's decisional processes. It would seem unlikely that decisions made by each individual member are independent. As opposed to parliaments, the EU Council gathers representatives of the governments of particular Union states, usually the ministers of a relevant field (depending on the subject matter being considered). Therefore, it may be expected that Council members will have common standards for evaluating the resolutions being considered.

How should the appropriate power index be selected?

Postulated power index properties

As we know, there exists a great variety of power indices. There is no single and universal power index, just as no universal and complete set of expectations for power indices has been established. This results from the diversity of preconditions for collective decision-making. Different postulates should be formed when dealing with decisions of a single house of parliament, where coalesced and stable political divisions play an extensive role. Our set of formal expectations would differ if we noted that the government coalitions being formed are exclusively minimal winning coalitions than if we had not made such an observation. Still other requirements may be formulated during analysis of complex decision-making processes, e.g. multi-house parliaments. The decision-making environment being the subject of this analysis is rather clearly defined. The postulates below account for some of its more important aspects.

The majority of votes may be presented as simple games in which the players can back or oppose a given proposal. The decision of the entire assembly is also dichotomous and involves either adopting or voting down a given proposal. A proposal may be passed only if supported by the majority defined in the decision-making rule (this is often a simple majority, and, at times, various qualified majorities of votes). We will consider the weighted voting systems (or weighted majority games) i.e. those where each voter is assigned a non-negative real number, i.e. the "weight," while a necessary and sufficient condition for adoption of the proposal is the voting coalition's attainment of the quota, i.e. the threshold established by the decision-making rule. Even complex voting systems may be presented as combinations of a number of weighted majority games.

A set of n players will be designated N and its particular constituents with small letters, i.e. $N = \{a_1, a_2, a_3, \ldots, a_n\}$. Function f assigns particular players their weights: $f(a_1)$, $f(a_2)$, $f(a_3)$ etc.

The quota will be designated q, where it may be expressed as the total number (total votes) or as percentage (share) of the total weight.

The weighted voting system, defined by two quantities, the quota and weight vector, will be expressed in the following form:

$$[q; f(a_1), f(a_2), f(a_3), \ldots, f(a_n)].$$

A subset of game participants will be referred to as a coalition and designated with capital letters, A, B, C, etc. Coalition C of game participants will be referred to as the winning coalition if the sum of the total weight of its members is equal to or greater than the quota required for the body to adopt the decision ($\sum_i f(a_i) \geq q$). Otherwise, the coalition will be a losing coalition. The set of all winning coalitions will be designated W.

Because a given coalition C is a winning coalition if and only if the total weight of its members is equal to or greater than q, then each coalition containing the winning coalition C is also a winning coalition. This particularly applies to the N set of all players.

A minimal winning coalition is a winning coalition, which becomes a losing coalition in the event of the defection of any player. The set of all minimal winning coalitions will be designated M (of course, set M is a subset of all winning coalitions W). The value of a certain power index for player a, i.e. the measure of its significance (power) in the decision-making collective will be referred to as $K(a)$.

What should be expected of the desired index? I will additionally attempt to present the political ramifications of each postulate.

Dummy player postulate

If, in the system under evaluation, a given player (voter, club, shareholder) is not necessary to make any coalition a winning coalition (adds no additional power to any coalition), the power index should be equal to 0 for such a player. This is identical to the absence of such player in any minimal winning coalition.

$$a \text{ – dummy player, i.e. } \sim(\exists C : C \in M \wedge a \in C) \Rightarrow K(a) = 0$$

This postulate is rather obvious and requires no further comment.

Symmetry (anonymity) postulate

The value of the power index should not depend on the order of the weights in the system. In other words, irrespective of how we record the system structure: $[q; f(a_1), f(a_2), f(a_3)]$, $[q; f(a_2), f(a_1), f(a_3)]$, $[q; f(a_2), f(a_3), f(a_1)]$ etc., we should attain the same set of winning and losing coalitions. In all

cases, we should obtain the same vector of power, which corresponds to the player vector permutation. The basis for the power index assigned to a player should only be the weights assigned thereto and the remaining players, and not their labels.

Despite differentiated relations among particular states, I decided to accept this postulate for two reasons. First, one of the fundamental assumptions accompanying the uniting of Europe is solidarity, thus behavior contrary to the divisions and aversion, which the Community is to "cover up." Second, even if this goal is not fully realized, the variety of decisions made over the coming years by the EU Council will reveal numerous axes of division among Union members. Therefore, over the long-term, it may be accepted that the symmetry among players will be maintained in establishing winning coalitions; each of the divisions (spatial dimensions) will be of equal significance.

Monotonicity postulate

If a certain player is weighted higher than another in a given system, that player's power should be greater than or equal to the latter, or $\forall\, i \neq j$: $(f(x_i) > f(x_j)) \Rightarrow (K(x_i) > K(x_j))$. Though this postulate is debatable for political reality teaches that a larger player is not always a desired coalition partner for the smaller player[4], this postulate seems necessary in selecting an appropriate analysis tool due to its intuitive nature. This is particularly important in describing complex voting systems, in which evaluation of the significance of particular system elements is not obvious ("if my state, e.g. state a has more votes than b, I would expect its position to be equal to or stronger than state b"). Last but not least, it is necessary to keep this (and the following) postulate for reasons of persuasion. The main audience for the analyses, i.e. politicians and their electorates, needs intuitive solutions.

Bloc postulate

If one of the players joining a block is a significant player, the power index of the bloc should be greater than the power index of the other player alone.

This postulate may be expressed as follows:

$$a, b \text{ – vote participants} \qquad (K(b) > 0) \Rightarrow (K(\{a\} \cup \{b\}) > K(a)).$$

Intuitively, this postulate may be presented as follows: it is difficult to imagine anyone joining powers with another who is needed in a coalition and

[4] See [9], [10], [21], [24].

not expect an increase in importance. Fulfilling this postulate will also be important during investigation of various coalitions' superadditivity.

Homogeneity assumption

The last assumption refers to a difference approach to power indices. Formulated by Straffin[5] and discussed in later works[6], it refers to the probabilistic interpretation of power indices. Straffin considers a voting model in which the probability that i-th voting for a resolution (p_i) may be treated as the value of a certain random variable that assumes values from a unit interval $[0, 1]$. Following that logic, we may ask: what is the probability that i-th's vote will impact the vote result? One of the proposals considered by Straffin[7] refers to the homogeneity assumption:

> For each i-th voter p_i is the same, equal number p as the value of a random variable distributed uniformly on the interval $[0, 1]$. (The homogeneity is among members: they all have the same probability p of voting for a given proposal, but p varies from proposal to proposal).

Straffin presents a valuable interpretation of this assumption. The probability p_i describes the level of acceptability for the i-th voter. Acceptance of the homogeneity assumption (for any voted matter $p_i = p$ is identical for all voters) means common evaluation standards for proposed resolutions to be voted on.

Shapley-Shubik Index

Of the most familiar and widely used indices to describe decision-making bodies, the only one, which fulfills these postulates, is the Shapley-Shubik index. It stands out among normalized power indices because it fulfills the transfer postulates[8]. Moreover, the above-referenced bloc postulate has a particularly valuable interpretation of said transfer postulates. Finally, this index defines the probability of impacting the voting result presuming acceptance of the homogeneity assumption.

Previous analyses usually referenced two indices, the Shapley-Shubik index and the Banzhaf index. The results obtained using both indices are

[5] [30], [31].

[6] See, *inter alia*, [4].

[7] To avoid an excess of topics, here, I will not discuss the proposed model in its entirety.

[8] See, inter alia, in works [4], [14], [15].

usually quite similar. However, opinions regarding the appropriate analysis tool are divided[9]. Acceptance of the above-referenced postulates made me apply the Shapley-Shubik index in my proposal.

This index also possesses an interpretation that properly recreates at least some qualities of the decision-making process in the Union Council. I mentioned above that with respect to the consensual character of Council decisions, the generally *ad hoc* constructed coalitions and lack of permanent agreements between Council members, negotiations will take numerous and differing paths, and decisions will be reached in varying ways over the long term. The Shapley-Shubik index assumes a random manner of consensus building (here, I am more akin to adopt equal probability of each way to attain consensus than of the equally probable all winning coalitions). Thus, it is referred to as the coalition building index.

New proposal – EU Constitution vs Nice Treaty

The European Convention appointed in 2001 drew up a draft European constitution. The proposed voting system was one of its most controversial points.

The table 2 presents the political and population weights of particular states after the Union enlargement.

The three-factor procedure provided for by the Nice Treaty has already been presented above. The Convention's initial draft introduced two conditions, which had to be met for the Union's Council to adopt a decision. This voting system dropped the weight criteria so key to the Nice system. Decisions must be backed by more than half the states with at least 60% of the European Union's population. Table 3 presents the criteria of both systems.

The ultimately accepted version of the voting system is based on the concept drafted by the Convention. It will be a two-factor system, in which **at least 55% of the states** (i.e. 15 out of 27) with **at least 65% of the European Union's population** will be needed to adopt resolutions put forward by the European Commission or EU Foreign Minister. These regulations are to be supplemented in 2009–2014, which in the politicians' opinion, **will secure the interest of the medium states against domination by large EU states**. These addenda are comparable to the Nice compromises (!) and provide that if a resolution proposed by the

[9] *Inter alia* [20], [22].

Table 2

Populations of member states as a % of EU population and number of votes (the "Nice weights")

State	Population (%)*	Number of votes
Germany	17.1	29
France	12.2	29
Italy	12.1	29
Great Britain	12.0	29
Spain	8.3	27
Poland	8.0	27
Romania	4.8	14
Holland	3.2	13
Greece	2.2	12
Czech Republic	2.2	12
Belgium	2.1	12
Hungary	2.1	12
Portugal	2.1	12
Sweden	1.8	10
Bulgaria	1.8	10
Austria	1.7	10
Slovakia	1.1	7
Denmark	1.1	7
Finland	1.1	7
Lithuania	0.8	7
Ireland	0.7	7
Latvia	0.5	4
Slovenia	0.4	4
Estonia	0.3	4
Cyprus	0.2	4
Luxembourg	0.1	4
Malta	0.1	3
Total	100.1**	345

* Data based on [3]
** The excess results from rounding

Table 3

Decisional rules according to the Nice Treaty and the Convention's initial draft

	Nice System	Constitutional System
Political weight criteria	258 votes (74,8%)	–
Population Criteria	62%	60%
Number of states criteria	Simple majority	Simple majority

European Commission is opposed by **at least 33.75% of EU states** (10 out of 27) or states populated by **at least 26.25% of the Union's population**, the Council's decision will be debated anew within a 'reasonable time'. Though the authors of the adopted draft maintain otherwise, the deferment for a 'reasonable time' is tantamount to blocking of the decision. After 2014, the EU Council will decide whether to remove this addendum.

Let us take a closer look at the superficially complex procedure presented above. First, a general remark. If opposition by a fraction of $1-q$ states or states populated by a percentage of $1-p$ Union residents suffices to block an adoption of a decision, then a fraction of states larger than q populated by a percentage p residents suffices to pass the same decision. For draft legislation presented by the EU Commission for an adoption by the Council, the support of more than 66.25% states with population comprising more than 73.75% of the Union's population is needed to push the legislation through. These are more stringent requirements than 55% of the states and 65%, respectively, of the population. Meeting those higher thresholds will be taken into account during negotiations since the inability to block a decision will mean its adoption by the body. Therefore, it is sufficient to account for only the said "addenda" in order to present the consequences of introducing such a supplemented system. We thus have a double majority system with a 66.25% threshold for states instead of the 55% and 73.75% threshold for population instead of 65%. The drawings below present the reference to division of power for the Nice system and the Convention's initial project to the division of power for the system established at the June 2004 EU summit.

It seems the figures 1 and 2 require no special comment. I will also eschew a detailed discussion of these and other proposals considered. You are quite familiar with these analyses especially since the readers and listeners of this work authored some of them.

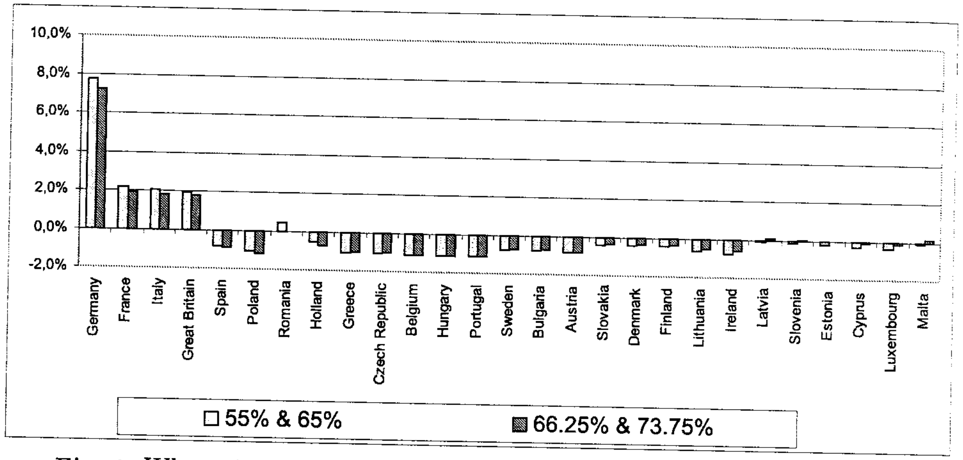


Fig. 1. Who gains and who loses on the introduction of the 55% & 65% and 66.25% & 73.75% system compared to the Nice system

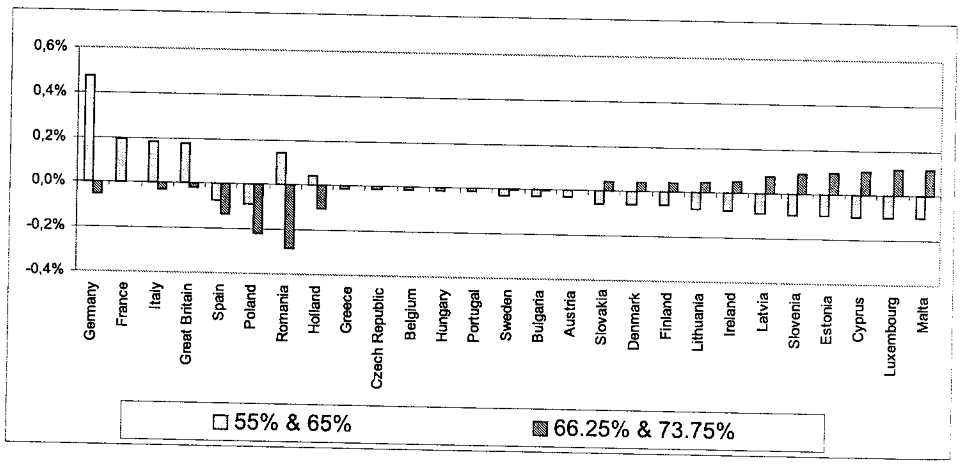


Fig. 2. Who gains and who loses on the introduction of the 55% & 65% and 66.25% & 73.75% system compared to the Convention's initial system

The fundamental issue I would like to discuss is resignation from the political weights, which were the basis for the previous EU Council voting systems as well as the Nice system. The consequences of this decision are very significant, not just for particular states (including my state, that was so "spoiled" by Nice), but for the future stability of a united Europe. This problem needs to be considered both in the context of the endurance of divisions within the European Union as well as in light of the Community's further enlargements.

Power of divisions in a united Europe

Investigation of the sub – as well as the superadditivity of various coalitions in a decision-making assembly allows for a projection of which voter coalitions will be treated thereby as valuable (depending on whether the coalition power is greater than the sum of the power indices of particular voters treated individually). Returning to the political interpretation of this property in the EU Council, we may attempt to answer the question of whether, in a given system, future votes will induce certain groups of states toward the conviction that it is worthwhile to vote in blocks (possibly establishing long-term strategies), or whether it would be more advantageous to vote separately without joining fixed coalitions.

Table 4

Superadditivity of selected coalitions in the basic EU Constitutional, minority blocking as well as in the Nice system

	Coalition	55 & 65	66.25 & 73.75	Nice
power of coalition	Germany & France	32%	30%	19%
total power of separated countries		27%	26%	17%
superadditivity		**118%**	**113%**	**111%**
power of coalition	Germany & Italy	32%	30%	19%
total power of separated countries		27%	26%	17%
superadditivity		**118%**	**114%**	**111%**
power of coalition	Germany & Great Britain	32%	30%	19%
total power of separated countries		27%	26%	17%
superadditivity		**118%**	**114%**	**111%**
power of coalition	France & Italy	25%	25%	19%
total power of separated countries		22%	21%	17%
superadditivity		**114%**	**120%**	**110%**
power of coalition	France & Great Britain	24%	25%	19%
total power of separated countries		21%	21%	17%
superadditivity		**114%**	**120%**	**110%**
power of coalition	Italy & Great Britain	24%	25%	19%
total power of separated countries		21%	21%	17%
superadditivity		**114%**	**118%**	**110%**

To eliminate an index's paradoxical properties, it is significant for the power index used to investigate superadditivity to fulfill the block postulate. The Shapley-Shubik index fulfills this postulate. I considered various hypothetical coalitions. One type of coalition includes states according to their size. In adopting a voting system that distinguishes states solely using population weights[10], we determine superadditivity of many coalitions built according to population size. The table 4 contains a comparison of dual-element coalitions established among four of the Union's largest states.

Clearly, the solutions adopted in June 2004 will provide a greater incentive for these states to vote in blocks than the Nice proposal did. In turn, this will influence a creation of countercoalitions among the states threatened by various "large tandems." This gives a rise to divisions. Increasing the durability of divisions seems to run counter to a Europe ideally founded on stability and collective construction of common values and market. Generally speaking, an organization in which divisions are reinforced has a greater propensity for deep crises.

Further European Union enlargement: What will EU states do when accepting Turkey into their ranks?

Politicians are discussing the realistic prospect of Turkey's accession to the European Union ever more frequently. The argument is that moderate Islam represented in Brussels would strengthen Europe's position with respect to other Islamic states, particularly respecting the ceaseless hotspots in the Middle East. Today, Turkey's population of 66 million would be the second only to Germany's. Each year, Turkey's population grows by about 1 million. According to demographic forecasts, by 2020, or ten years after implementation of the new voting system, Turkey will be the most populous state with territory on the Old Continent. Turkey's vote will then be more important than that of the Union's largest founders, which have far more extensive economic potential and rightly expect the leading position in the Community's decision-making process.

The question posed above merely indicates the most extreme case exemplifying a more general problem. The complete abandonment of politicized

[10] Each of the proposed two-factor systems was a conjunction of two weighted majority games: games with weights identical for each country and a game with the weights being fractions of particular states' populations. They differed only in their threshold levels.

weights and basing of the decision-making procedure on population weights will, in the future, cause Europe to feverishly seek additional solutions to maintain a balanced process. These will clearly be arbitrary solutions that call upon strictly political arguments. Nonetheless, this can be foreseen right now. The only way to avoid sketchy modifications of the voting system was to adopt negotiable and arbitrary political weights (not necessarily of the Nice variety). When building a community, it is impossible to avoid negotiations, even ones as prickly as "which of us is the most important?" Therefore, paradoxically, the maintenance of political weights will serve the stability of the future European Union and preserve the influence of its undisputed leaders.

How to reach consensus?

Numerous proposals have been formulated as the basis for compromise. Contrary to politicians' public declarations regarding the need to increase the system's effectiveness[11] and equality among the vote of the European Union's citizens, a system with lower effectiveness than proposals submitted during negotiations, which does not equalize the vote at all, was ultimately selected. This occurred, even though there were systems proposed that referred to effectiveness or those that postulated equality of the vote[12]. Thus, other arguments convinced the politicians. On what basis was one solution selected over another? When comparing the balance of power in the selected system, in the Convention's proposal and the Nice system, it seems the most important issue was the need to reinforce the position of the Union's biggest states. However, it remains uncertain whether the rather intricate system selected will actually meet the negotiators' expectations. The agreements give the impression of being random and resulting from "groping in the dark." The indictments against the adopted formula[13] (including those mentioned above) leave little room for optimism.

Below, I would like to present a method that would satisfy the formal and political postulates in establishing a voting system.

[11] With respect to the consensual nature of the agreements in the Council this postulate should not be regarded as fundamental.

[12] See, *inter alia*, [1], [5], [12], [13], [22], [26], [27].

[13] [7], [12], [28], [29].

Compromise – description of the method

1. Single majority system. The multifactor voting systems considered by politicians are unnecessary. It can be shown that the Nice system and the two factor systems have corresponding forms in weighted voting systems (with a single weight criteria – the quota), equivalent with respect to the division of voters' power. Postulating a simple system postulate would provide a clear and transparent system that is easier to present to public opinion.

2. Political weights. How should the demographic, economic and political differences of individual European Union states be accounted for? There is no way to avoid negotiations. Here, it makes sense to discuss the general interpretation of power indices. Power values assigned by indices to particular players actually correspond to certain attitudes of decision-making bodies' members that arise from experiences collected over a large number of votes. They express the conviction of the body's members regarding the need for them and their partners in the voting process. The power index values are a consequence of the vote configuration in a decision-making body, "researched" numerous times by decision makers when establishing coalitions in connection with votes. Thus, it may be said that the power index values are the magnitude that is first and foremost directly felt by the vote participants, rather than the weights. Previous negotiations concerned construction of the voting system and establishing **weights** satisfactory to all. In fact, however, the dispute concerned the **significance** of particular states and such were the arguments raised. Consequently, the negotiations were not topical. In order to sketch out the appropriate arrangement of weights-votes in the assembly, the opposite of what had been previously done needs to be done now. Ask the politicians themselves regarding their expectations – not expectations regarding the allocation of votes – but, rather, concerning the importance of their states in the decision-making process in matters of significance to the European Union. Through negotiations, the politicians would establish the proportional powers of particular states in the Council of the European Union by allocating successive fragments of the total "power pie". With a certain value arrangement for a given index of power at hand, it is then possible to precisely designate a correspondingly weighted voting system. Accordingly, with an established decision-making rule (weight criteria – quota) it is possible to find a weight arrangement for which the allocation of power will be the same or very near as that which has been assumed (with 27 players it can be accepted that the allocation of power implied by the weight arrangement would be practically identical with that which has been assumed). Finding such an arrangement in

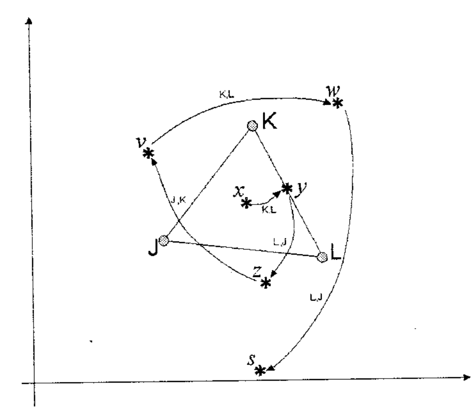
practice is not entirely simple but doable. However, this would not be an indisputable composition – a greater number of such weight arrangements may exist.

3. Decision-making rule. An issue to be yet discussed during the analysis of the decision process is the problem of vote result instability referred to in the voting theory as voting cycles. The occurrence of cycles creates a possibility for strategic manipulation. If decisions are made in the collective body sequentially (in a successive vote, the winning alternative competes with the next on the agenda until the exhaustion of a set of alternatives) the resulting arrangement may be dependant on the order of resolutions presented for voting. This property is referred to as the voting paradox (or Condorcet's paradox).

Due to the multidimensionality of decisions taken in the European Union Council, the spatial aspect of the voting cycles problem must be accounted for. The problem then becomes even more complex, as indicated by McKelvey's theorem[14]. It is not sufficient to assume that decision makers' preferences are unimodal. Excepting situations of exceptional ideal points' symmetry, there is no Condorcet's winner. The example below illustrates the problem:

Example 1

Let us consider three voters, J, K, L with ideal points in two-dimensional space, as in the figure below.



Source: [8]

Fig. 3. McKelvey's Catapult

14 [17], [18].

Even though option X seems advantageous to the entire voting trio (it is located within the triangle mapped out by their ideal points, thus being as "near" as possible to the entire group – it belongs to the Pareto set), the coalition of K and L, who would prefer alternative Y, may vote against it. That, however, would lose to alternative Z supported by the coalition of J and L. Alternative Z would lose to V, which would be beaten by alternative W, to be in turn beaten by S (the letters next to the arrows indicate the coalition, which is interested in backing a given change). Thus, a party controlling the voting agenda could instigate selection of any option. This is not a matter only for voting theorists. It may turn out to be a tremendous headache for politicians. On the one hand, a seasoned leader of an assembly familiar with voters' preferences and the extent of the political dispute, may successfully shepherd along certain resolutions favorable to him or her. On the other hand, even the genuine and sincerely expressed will of the body's members (e.g. the EU Council) may be subject to accusations of manipulation. This is especially so in the recently expanded composition.

In 1988, Caplin and Nalebuff[15] proved a theorem, which may be treated as one of the best proposals for resolving the problems presented above. They showed that if voters' preferences are unimodal and if their positions in the collective body are not extremely polarized (an assumption regarding the arrangement of voters' ideal points being concave[16]), then application of the qualified majority rule of no less than 64% (precisely $1 - 1/e \approx 0,632$) in a voting arrangement will prevent the appearance of cycles. It will consequently assure stability of the assembly's decisions. Recall that e.g. in parliaments, issues of particular importance require sizeable qualified majorities, such as a 2/3 vote in favor.

We are thus familiar with the sought value of the qualified majority of 64%. Higher values would decrease the collective's efficiency. I do not consider this to be a key issue, but it is certainly the one to be kept in mind.

4. Establishing weights. Above, I wrote that the established allocation of power and quota do not determine allocation of weights. With an established decisional rule, it is possible to find various weight arrangements, which would result in precisely the same coalitions being formed.

15 [2]

16 This condition is met if $p[\lambda f(W_s) + (1-\lambda)f(W_t)] \geq \lambda p[f(W_s)] + (1-\lambda)p[f(W_t)]$, where $p[f(W_s)]$ and $p[f(W_t)]$ designate the density of voters' ideal points, respectively, W_s and W_t, while $0 \leq \lambda \leq 1$.

The method for selecting the most appropriate one among formally equivalent weight arrangements is actually an open question. It may be served by the **transparency criterion** proposed by Słomczyński and Życzkowski[17]. According to that criterion, the selected weight allocation should minimize the deviation from the power allocation. Traditional statistical parameters may be accepted to measure the extent of said deviation (but this is another matter).

The advantages of the suggested manner of designing a decision-making system in the EU Council include stable relations among states. In the event of successive Union enlargements, negotiated power proportions would remain unchanged and would not require additional negotiations. It would be only necessary to establish the relative position of the state being accepted into the Community. After establishing the power proportions of the new Union member with respect to the power index values of the "old" members, it would then suffice to convert the values of all power indices to total 100%. Meanwhile, the procedure for adapting the weights to the new power allocation (using the quota of 64%) would remain unchanged.

Example 2

To simplify the matter, we will consider a three-person collective, in which voter A, pursuant to prior arrangements among the voters, has four times more power than C and twice that of B. The power allocation will be $[57\frac{1}{7}\%, 28\frac{4}{7}\%, 14\frac{2}{7}\%]$. Let us assume the newly arrived member, say D, would have more power than B but less than A. We will accept the following negotiated power proportions for voters A, B, C and D: $4:2:1:3$. Therefore, the new allocation of power in collective $\{A, B, C, D\}$ would be as follows $[40\%; 20\%; 10\%; 30\%]$. It is sufficient to solve a simple equation with one variable. It is easy enough to verify that the initial and negotiated proportions have been maintained.

Clearly, the allocation of power will "flatten out" with the accession of successive members if the previously established proportions are maintained. This is a natural consequence of divvying up the same "pie" (i.e. 100%) into ever-smaller pieces. However, this procedure guarantees the stability of power proportions in the collective body.

17 [27]

Summary

The trials of negotiation are unavoidable on the road to finding an appropriate voting system for politicians. What is vital, however, is that these negotiations concern the correct topic, i.e. the actual significance of member states, not the weights assigned thereto. By following the procedure proposed herein, it would be easier to achieve actual compromise. What is more, the proposal presented would provide increased opportunity for a genuine debate, where technical details regarding the manner of decision-making would not interfere with the intentions of negotiators. If an understanding is reached as to the proportion of power, realization of its technical aspects (establishing the weights) would be a task for scientists, not politicians and their bureaucrats.

References

[1] Bilbao, Jesus M., Julio R. Fernandez, Nieves Jimenes and Jorge J. Lopez. 2002. *Voting power in the European Union enlargement.* "European Journal of Operational Research" 143: 181–196.

[2] Caplin, Andrew, Barry Nalebuff. 1988. *On 64%-majority rule.* "Econometrica" 56(4): 787–814.

[3] European Center Natolin. 2003. *Costs and Benefits of Poland's Membership in the European Union.* Warsaw.

[4] Falsenthal, Dan S. i Moshé Machover. 1998. *The Measurement of Voting Power.* Cheltenham: Edward Elgar.

[5] Falsenthal, Dan S. i Moshé Machover. 2001. *The Treaty of Nice and Qualified Majority Voting.* "Social Choice and Welfare" 18: 431–464.

[6] Falsenthal, Dan S. i Moshé Machover. 2003. *Analysis of QM rules in the draft Constitution for Europe proposed by European Convention.*

[7] Falsenthal, Dan S. i Moshé Machover. 2004. *Analysis of QM Rule Adopted by the EU Inter-Governmental Conference Brussels, 18 June 2004.*

[8] Haman, Jacek. 2003. *Stabilność i zmienność w przestrzennych modelach głosowania.* "Studia Socjologiczne" 1: 39–78.

[9] Jasiński, Mikołaj. 2000. *Czy zawsze większy jest silniejszy, czyli jak zmierzyć siłę uczestników zgromadzeń decyzyjnych?* "Studia Socjologiczne" 1–2: 49–77.

[10] Jasiński, Mikołaj. 2003. *Stanowisko ideologiczne a znaczenie uczestnika zgromadzenia decyzyjnego.* "Studia Socjologiczne" 1: 139–174.

[11] Kirsch, Werner. 2004. *The new qualified majority in the Council of the EU. Some comments on the decisions of the Brussels summit.* Bochum: preprint.

[12] Lane, Jan-Erik, Reinert Maeland. 2002. *Research note. A note of Nice.* "Journal of Theoretical Politics" 14(1): 123–128.

[13] Mabille, Lionel. 2003. see: N. Dubois, *Pondération des voix: la preuve par "27".* Liberation, 26/11/2003.

[14] Malawski, Marcin. 2001. *Własności transferu indeksów siły w schematach decyzyjnych.* Warszawa: IPI PAN, conference material.

[15] Malawski, Marcin. 2002. *Equity properties of the Shapley value as a power index.* "Control and Cybernetics" 31.

[16] Malawski, Marcin, Honorata Sosnowska and Andrzej Wieczorek. 1997. *Konkurencja i kooperacja. Teoria gier w ekonomii i naukach społecznych.* Warszawa: PWN.

[17] McKelvey, Richard. 1976. *Intransitivities in multidimentional voting bodies and some implications for the agenda control.* "Journal of Economic Theory" 12: 472–482.

[18] McKelvey, Richard. 1979. *General conditions for global intransitivities in formal voting models.* "Econometrica" 47(5): 1085–1112.

[19] Michałowska-Gorywoda, Krystyna. 2002. *Podejmowanie decyzji w Unii Europejskiej.* Warszawa: WN "Scholar".

[20] Nurmi, Hannu, Tommi Meskanen and Antti Pajala. 2001. *Calculus of Consent in the EU Council of Ministers.* In: Power indices and Coalitional Formation, M. J. Holler and G. Owen (ed.) Boston: Kluwer Academic Publishers: 291–313.

[21] Owen, Guillermo. 1971. *Political games.* "Naval Research Logistics Quarterly" 18: 345–355.

[22] Paterson, Iain, Peter Silárszky. 2003. *Draft Constitution and the IGC: Voting can be simple and efficient – without introducing the massive transfer of power implied by the Convention's double majority proposal.* Short Policy Paper for EU Convention Forum No. 2. Institute for Advanced Studies, Vienna.

[23] Shapley, Lloyd S. 1953. *A value for n-person games.* In: Contributions to the Theory of Games. v. 2. Princeton: Princeton University Press: 307–317.

[24] Shapley, Lloyd S. 1977. *A Comparison of Power Indices and a non--Symmetric Generalization.* RAND Paper. Santa Monica: Rand Corporation. P-5872.

[25] Shapley, Lloyd S., Martin Shubik. 1954. *A method of evaluating the distribution of power in a committee system.* "American Political Science Review" 48: 787–792.
[26] Słomczyński, Wojciech, Karol Życzkowski. 2004a. *Wyliczony kompromis.* "Rzeczpospolita" 25: 30 I.
[27] Słomczyński, Wojciech, Karol Życzkowski. 2004b. *Rules Governing Voting in EU Council: The Treaty of Nice, the Draft Treaty Establishing a Constitution for Europe and a Proposed Compromise Solution.* Kraków: preprint.
[28] Słomczyński, Wojciech, Karol Życzkowski. 2004c. *Traktat z Nicei, projekt Konwentu i Konstytucja Europejska. Porównanie mechanizmów decyzyjnych w Radzie Unii Europejskiej.* Kraków: preprint.
[29] Sozański, Tadeusz. 2004. *Kompromis?* Kraków: preprint.
[30] Straffin, Philip D. 1982. *Power Indices in Politics.* In: Political and related models, S. J. Brams, W. Lucas i P. D. Straffin (ed.) New York: Springer Verlag: 256–321.
[31] Straffin, Philip D. 1994. *Power and Stability in Politics.* In: Handbook of Game Theory, v. 2, R. J. Aumann and S. Hart (ed.) Amsterdam: Elsevier Science: 1127–1151.
[32] Styczeń, Marek. 2003. *Computer program "Nicea".*

Mikołaj Jasiński
Department of Statistics, Demography and Mathematical Sociology
Institute of Sociology
Warsaw University
e-mail: mikolaj.jasinski@uw.edu.pl

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 7 (20) 2004

Andrzej Malec
The University of Bialystok

SOME LOGICAL ASPECTS OF THE ACCESSION OF POLAND TO THE EUROPEAN COMMUNITIES

Summary. The paper has two main aims. The first aim is to put forward an explanation of the principle of superiority of the law of the European Communities (*acquis communautaire*) in relation to the system of Polish domestic law in terms of rules of legal reasoning. The second aim is to consider the validity of the above principle in terms of an analogy between the notion of the system of law and the notion of a deductive system.

Introduction

It is generally accepted by Polish lawyers that from the moment of Polish accession to the European Communities, they have to obey the principle of superiority of the law of the European Communities (*acquis communautaire*) in relation to the system of Polish domestic law. This principle can be incorporated into the system of Polish domestic law by introducing some new rules of legal reasoning. These new rules of legal reasoning can be divided into three subsets:

- rules of semiotic interpretation,
- rules of functional interpretation

and

- collision rules.

New rules of semiotic interpretation

The rules of semiotic interpretation allow us to derive legal norms from a legal text using just semiotic properties of the text[1]. The set of such rules contains, for example, the following rules:

[1] A legal text is a set of inscriptions given by the lawmaker as a source of law. The Civil Code of Poland is an example of a legal text. The lawmaker is a fictitious person recognized as an author of all legal texts.

ISBN 83-7431-015-4 ISSN 0860-150X

- it is forbidden to omit any fragment of a legal text during the interpretation of that text,
- it is forbidden to add any phrases to a legal text during the interpretation of that text,
- it is an obligation of the interpreter to understand any expression W of a legal text in accordance with a legal definition of this expression (the legal definition of an expression W is a definition of this expression contained in a legal text[2]).

The principle of superiority of *acquis communautaire* can be expressed by the following rules of semiotic interpretation:

- the rule of preference of legal definitions *sensu stricto* contained in the legal texts of the European Communities,
- the rule of preference of legal definitions *sensu largo* contained in the legal texts of the European Communities,
- the rule of preference of the legal language of *acquis communautaire.*

According to the rule of preference of legal definitions *sensu stricto* contained in the legal texts of the European Communities, if:

- the meaning of an expression W from a Polish legal text T is not sufficiently clear for the purpose of the interpreter of that legal text

and

- the text T is an implementation of the community law,

it is an obligation of the interpreter to make clear the meaning of the expression W in accordance with the legal definitions *sensu stricto* of the expression W contained in the legal texts of the European Communities, unless there are special reasons for ignoring these definitions.

According to the rule of preference of legal definitions *sensu largo* contained in the legal texts of the European Communities, if:

- the interpreter has used the rule of preference of legal definitions *sensu stricto* contained in the legal texts of the European Communities

and after that

- the meaning of an expression W from a Polish legal text T remains not sufficiently clear for the purpose of the interpreter of that legal text,

it is an obligation of the interpreter to make clear the meaning of the expression W in accordance with the legal definitions *sensu largo* contained in the

[2] The legal definitions can be divided into two subsets: legal definitions *sensu stricto* and legal definitions *sensu largo*. The legal definition *sensu stricto* of an expression W is placed in a legal text by the lawmaker with the intention to establish or to make clear the meaning of the expression W. The legal definition *sensu largo* of an expression W is a part of a legal text which determines the meaning of the expression W and it was placed in the legal text by the lawmaker with the intention to establish a legal norm.

legal texts of the European Communities, unless there are special reasons for ignoring these definitions.

According to the rule of preference of the legal language of *acquis communautaire*, if:

- the interpreter has used the rule of preference of legal definitions *sensu stricto* contained in the legal texts of the European Communities

and

- the interpreter has used the rule of preference of legal definitions *sensu largo* contained in the legal texts of the European Communities

and after that

- the meaning of an expression W from a Polish legal text T remains not sufficiently clear for the purpose of the interpreter of that legal text,

it is an obligation of the interpreter to make clear the meaning of the expression W in accordance with:

- the interpretations of the expression W given by the Courts of the European Communities

and

- the interpretations of the expression W given by EC lawyers,

if these interpretations were generally accepted before the date of issue of the legal text which is the subject of interpretation.

New rules of functional interpretation

The rules of functional interpretation allow us to derive legal norms from a legal text whenever rules of semiotic interpretation are insufficient or inappropriate to our task. The rules of functional interpretation are mainly based on valuation of aims, goals and intentions of the lawmaker. The set of such rules contains, for example, the following rules:

- it is forbidden to accept an interpretation of a legal text which contradicts the intentions explicitly expressed by the lawmaker,
- if the meaning of an expression W is not sufficiently clear after the process of semiotic interpretation, it is an obligation of the interpreter to make the meaning of the expression W clear in accordance with aims, goals and intentions expressed by the lawmaker.

The principle of superiority of *acquis communautaire* can be expressed by the following rules of functional interpretation:

- the weak rule of preference of the values of *acquis communautaire*,
- the strong rule of preference of the values of *acquis communautaire*.

According to the weak rule of preference of the values of *acquis communautaire*, if:

- the interpreter has used the rules of semiotic interpretation

and after that

- the meaning of an expression W from a Polish legal text T remains not sufficiently clear for the purpose of the interpreter of that legal text,

it is an obligation of the interpreter to make the meaning of the expression W clear in such a way that the norms derived from the text T containing the expression W are coherent with the system of generally accepted values of *acquis communautaire*. The system of generally accepted values of *acquis communautaire* can be reconstructed from interpretations given by the Courts of the European Communities and interpretations given by EC lawyers, even if these interpretations were not generally accepted before the date of issue of the legal text which is the subject of interpretation.

According to the strong rule of preference of the values of *acquis communautaire*, if:

- the interpreter has used the rules of semiotic interpretation

and after that

- the meaning of an expression W from a Polish legal text T is sufficiently clear for the purpose of the interpreter of that legal text,

however,

- the meaning of an expression W from a Polish legal text T is inappropriate from the point of view of EC values

it is an obligation of the interpreter to make the meaning of the expression W clear in such a way that the norms derived from the text T containing the expression W are coherent with the system of generally accepted values of *acquis communautaire*.

A new collision rule

The rules of collision are used to solve collisions of legal norms. The set of such rules contains, for example, the following rules:

- *lex posterior derogat legi priori* (later norms suppress earlier norms),
- *lex superior derogat legi inferiori* (superior norms suppress inferior norms),
- *lex specialis derogat legi generali* (particular norms suppress general norms).

The principle of superiority of *acquis communautaire* as a collision rule can be expressed in the following way: it is an obligation of the interpreter to ignore a norm derived from a Polish legal text by either the rules of semiotic interpretation or the rules of functional interpretation whenever this norm contradicts any norm of *acquis communautaire* (as a rule – a norm

of *acquis communautaire* contained in an European directive or an European regulation).

The system of law as a deductive system

It is possible to treat any system of law of continental Europe as an analogue of a deductive system:

- legal texts constitute the set of its axioms

and

- the generally accepted rules of legal reasoning constitute the set of its rules of inference[3].

It is obvious for all logicians that the axiomatic form of a deductive system can be replaced by the form without any axioms (the form with rules of inference only). And vice-versa: the form of a deductive system without any axioms can be replaced by the axiomatic form. Let us look from that point of view at the system of Polish law. Before the accession it was constituted by the set of legal texts **T** and the set of generally accepted rules of legal reasoning **R**. After the accession we added a set **ET** of all legal texts of the European Communities to the set **T**. Also we added the above six rules of legal reasoning to the set **R**.

The following question can be considered:

- is the principle of superiority of *acquis communautaire*, expressed by the above six rules of legal reasoning, expressible in terms merely of the set **R** and the new set of valid legal texts **T** + **ET** (texts which are valid from the date of accession)?

In other words:

- can the principle of superiority of *acquis communautaire* be derived from the set of legal texts **T** + **ET** by former rules of legal reasoning (the rules from the set **R**)?

The principle of superiority of *acquis communautaire* vs. the constitution of Poland

First of all, the principle of superiority of *acquis communautaire* is not explicitly given in any legal text of the European Communities. It means that this principle cannot be derived from the set **ET** of legal texts of the European Communities only by the rules of **semiotic interpretation**.

[3] Of course, it is not correct to say about any real system of law that this system is a deductive system. We are just talking about some possible **analogy**.

However, the above principle was derived from the set of legal texts of the European Communities. It was derived, thanks to the rules of **functional interpretation**, by the European Court of Justice, which has a power to interpret the community law. So, the European Court of Justice assumed that the principle of superiority of *acquis communautaire* is a part of *acquis communautaire*. In fact, the rules of functional interpretation which were used by the European Court of Justice are not elements of **R**, because they are related to the concept of effectiveness of the community law.

Thus the principle of superiority of *acquis communautaire* cannot be derived from the set of legal texts **ET** by the rules from the set **R**.

Moreover, even if we assume for a moment that it is possible to derive the principle of superiority of *acquis communautaire* from the set of legal texts **ET** by the rules from the set **R**, we will find one important problem: from the set of legal texts **T** by the rules from the set **R** a principle can be derived which contradicts the principle of superiority of *acquis communautaire*. The act of accession is just an international contract and from the point of view of our constitution it is placed below the constitution in our hierarchy of legal texts, however, it is placed higher then all other legal texts except the constitution.

Therefore, the principle of superiority of *acquis communautaire* cannot be derived from the set of legal texts **T** + **ET** by the rules from the set **R**.

Andrzej Malec
Department of Logic, Computer Sciences
and the Methodology of Sciences
The University of Bialystok
malec@hum.uwb.edu.pl

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 7 (20) 2004

Renata Jermołowicz
Technical University of Białystok

REMARKS ON ADAM SMITH'S LECTURES ON RHETORIC AND BELLES LETTRES

A lawyer without history or literature is a mechanic, a mere working mason; if he possesses some knowledge of these, he may venture to call himself an architect...

Sir Walter Scott

Teach you children poetry; it opens the mind, lends grace to wisdom and makes the heroic virtues hereditary...

Sir Walter Scott

Until recently the fame of Adam Smith rested mainly upon his revolutionary work entitled *The Wealth of Nations*. Published in 1776, the year of the American Revolution, it remains one of the classical textbooks of the history of economics where Smith establishes economics as an autonomous subject and introduces the doctrine of free enterprise. Though today the figure of Adam Smith is primarily associated with the history of economics, it is also important to realize that he was not merely an economist.[1] It should be remembered that *The Wealth of Nations* was the embodiment of Smith's wide interests, embracing not only economics, ethics, political philosophy, and jurisprudence, but also history of science, psychology, and literature. A quite recent discovery of students' notes on Dr Smith's lectures on rhetoric and belle lettres reveal not only his genuine ease with literature (both ancient and modern) and the art of writing but also his role in the transition from the earlier, well-established tradition of formal rhetoric at the Scottish universities.

[1] Smith himself regarded his *Wealth of Nations* only as a partial exposition of a much larger work where he hoped to display general principles of law and government and different revolutions they have undergone in the course of history. Unfortunately, Smith never completed it in his lifetime.

ISBN 83-7431-015-4 ISSN 0860-150X

In 1958 the remnants of the village library of Whitehaugh were displayed on auction in Aberdeen. Professor John M. Lothian was lucky to be among the purchasers. The library contained many valuable volumes among which there happened to be the first editions of Thomas Hobbes, Thomas Reid and Adam Smith.[2] Among the numerous manuscripts displayed at the Aberdeen sales there were two sets of notes taken by students. One of them was a course on Jurisprudence delivered by Adam Smith. The second set was entitled "Dr Smith's notes on Rhetorick Lectures" and also appeared to be a complete set of notes taken on part of Smith's lectures on Moral Philosophy in Edinburgh. Professor Lothian purchased the notes and edited them in 1958. W. R. Scott, a devoted bibliographer of Adam Smith, comments on the lectures: "There is a certain amount of mystery about the circumstances and the subject-matter of the lectures which Adam Smith delivered at Edinburgh during the three years 1748–1751."[3] Professor Lothian, who wrote the introduction to the *Lectures on Rhetoric and Belle Lettres*, took the pain of discovering the circumstances of the Edinburgh lectures. He claims that after his return to Edinburgh in 1746 Smith was in need of a post which would ensure him permanent financial benefits. Apparently it was due to Smith's friends (among whom there were Lord Kames, James Oswald, and Robert Craigie) that Smith could earn some money by lecturing.[4] They came up with the idea of lectures on rhetoric and belles lettres which would be delivered by Smith and which would be completely different from any available at the time.

While discussing the circumstances of the Edinburgh lectures on rhetoric and belle lettres it is necessary to remember that Smith's lectures coincided with a great intellectual awakening in Scotland. Literature and rhetoric had become a fashionable hobby among the noblemen who hired tutors to give lectures on literature and taste at their homes. Numerous intellectual clubs and societies had been organised. The members of those clubs and societies aimed at propagating literature and the arts of speaking and writing. They held various competitions to encourage both professionals and amateurs to submit essays on a given subject. The best essays were read and discussed at the societies' meetings. Smith was a member of the famous Select Society whose goal was "philosophical inquiry, and the

[2] J. M. Lothian (ed.), *Lectures on Rhetoric and Belles Lettres delivered in the University of Glasgow by Adam Smith reported by a student in 1762–63*, Thomas Nelson and Sons Ltd, Toronto and New York, 1963, p. XII.

[3] W. R. Scott, *Adam Smith as Student and Professor*, Glasgow, 1973, p. 46.

[4] J. M. Lothian, *op. cit.*, p. XIII.

improvement of the members in the art of speaking."[5] In 1755 The Select Society announced a competition for the best essay on Taste.[6] The interest in writing provoked a quick development of newspapers and periodicals – "Scots Magazine", "Tatler", "Rambler" and "Edinburgh Review" – which were a popular way to encourage the understanding of literature and philosophy. Smith spoke highly of the Scottish periodicals and even wrote a few articles which he contributed anonymously to "Edinburgh Review".[7]

Therefore, it was in the atmosphere of public interest in literary topics that Smith grasped at the prospect of delivering novel lectures on rhetoric and belle letters. Indeed, he was well-equipped to introduce a spirit of novelty to the dull classical school of rhetoric which he regarded as "a very silly set of books not at all instructive".[8] At Oxford Smith himself enjoyed a thorough studies of Greek and Latin literature, history and philosophy. According to his contemporaries, Smith's knowledge of Greek and Latin literature was uncommonly accurate and extensive.[9] He was also known for his remarkable ability to call to mind long passages of Latin and Greek authors. His students were amused with his eloquence: "Those who receives instruction from Dr Smith will recollect with much satisfaction many of those incidental and digressive illustrations and discussions, not only in morality but in criticism, which were delivered by him with animated and extemporaneous eloquence as they were suggested in the course of question and answer."[10] Smith adored literature – especially he liked Swift, Dryden, Pope and Gray whose style of writing he regarded as clear and precise. Moreover, he was at ease with English, French and Italian writers and often translated passages from French and Italian in order to improve his style. He used translation as a part of the lectures encouraging his students to translate passages of their favourite writers in order to analyse and compare the niceties of grammar.[11] Certainly, Smith, with his love for literature, was the right person to launch a new attitude towards lectures on rhetoric. The lectures brought him a great success: his reputation as a professor was

[5] J. M. Lothian, *op. cit.*, p. XXIV.

[6] David Hume acted as a member of the jury. The jury chose the best essay which was written by Professor Gerard – a famous figure in Aberdeen.

[7] One of his best articles was the analysis of Johnson's Dictionary of the English Language.

[8] A. Smith, in J. M. Lothian (ed.), *op. cit.*, *Lecture 6.*

[9] J. Rae, *Life of Adam Smith*, London, 1985, p. 23.

[10] J. Rae, *op. cit.*, p. 56.

[11] Smith regarded translation as an art; he often criticized a word-by-word literary translation which, according to him, becomes ambiguous and loses the very essence of the author's intention.

very high, and in a short time a multitude of students from far away moved to Edinburgh to participate in the lectures given by Dr Smith.

Thus, the students were presented with a quite different content of lectures on rhetoric enriched with the analysis of belle lettres and the art of writing. John Millar, one of the most beloved students of Smith, commented on the content of the lectures: "he (Dr Smith) saw the necessity of departing widely from the plan that had been followed by his predecessors, and of directing the attention of his pupils to the studies of a more interesting and useful nature than the logic and metaphysics of the schools".[12] Smith saw the opportunity to conduct his students towards the power of reasoning through stimulating their feelings and aesthetic sense: "The best method of explaining and illustrating the various powers of the human mind, the most useful part of metaphysics, arises from an examination of the several ways of communicating our thoughts by speech, and from an attention to the principles of those literary compositions which contribute to persuasions or entertainment. By these arts, everything that we perceive and feel, every operation of our minds, is expressed in such a manner, that it may be clearly distinguished and remembered. There is, at the same time, no branch of literature more suited to youth at their first entrance upon philosophy than this, which lays hold of their taste and their feelings".[13]

Therefore, during the thirty lectures on rhetoric and belle letters students were taught to analyse literary works in order to understand the principles of good composition with the emphasis on narrative and descriptive prose. Smith began by discussing the origins of language; he compared the language to a primitive machine which gained its complexity as it developed throughout the centuries.[14] Since the rules governing the spoken and written language had undergone transformation, Smith intended to analyse the evolution of the written and spoken word with the emphasis on different manners of describing events, characters and scenery in historical, narrative and judicial writing. The ideal was to lead students towards perfection of style which, in his opinion, consisted in "expressing in the most concise, proper, and precise manner the thought of the author, and that is in the manner which best conveys the sentiment, passion, or affection with which it affects – or he pretends it does affect – him, and which he designs to communicate to his reader. This, you'll say, is no more than common sense:

[12] D. Stewart (ed.), A. Smith, *Theory of Moral Sentiments*, London, 1983, p. XVI.
[13] D. Stewart, *op. cit.*, p. XVI.
[14] A. Smith, in J. M. Lothian (ed.), *op. cit.*, *Lecture 3.*

and indeed it is no more."[15] Interesting is the fact that the appeal to write in a simple, clear style appears almost in every lecture.

Before Smith's students proceeded to a more practical part of the lectures that is writing, they enjoyed a thorough comparison and analyses of various styles and manners – the styles of Herodotus, Thucydides, Xenophon, Polybius, Livy, Machiavelli, and, of course, Smith's favourite Swift were tackled. With the help of numerous literary and historical examples students were taught how to select the right events to include and how to treat their causes, both proximate and remote. Smith's ease with the history and historians allowed his students to follow the evolution of historical descriptions. From historical writing Smith proceeded to expository writing. At that stage students were taught to differentiate between the elements contributing to variety, unity and decorum of the writing text. Here Aristotelian didactic discourses and Ciceronian deliberative orations were discussed. At the next stage of the lectures Smith discussed the nature of judicial oratory of Greece and Rome and contrasted it with the English school of oratory. Again, Smith emphasized the importance of "a natural order of expression, free of parentheses and superfluous words."[16] Finally, students were challenged to put their knowledge of good composition in practice and write an elaboration on a given subject. Thus, Smith's attitude towards the content of the lectures had much of a practical nature – his primary goal was to teach them to write in a simple and direct style without much decorum. Here and there Smith warns against the abuse of metaphors, allegories, similies, metonymies, and hyperboles which, when overused, may lead to ambiguity.[17]

While discussing the phenomenon of Smith's lectures it is impossible to undergrade his personal qualities. His strong, analytical mind mixed with his passion for literature created the atmosphere of creative scientific research and students could hardly be bored on his lectures. John Millar commented on Smith's manner: "In delivering his lectures, he trusted almost entirely to extemporary elocution. His manner, though not graceful, was plain and affected; and, as he seemed to be always interested in the subject, he never failed to interest his hearers. [...] even the small peculiarities in his pronunciation or manner of speaking became frequently the objects of imitation."[18] It was the persona of Dr Smith that attracted the gentlemen whose reputa-

[15] A. Smith, in J. M. Lothian (ed.), *op. cit.*, *Lecture 11.*
[16] A. Smith, in J. M. Lothian (ed.), *op. cit.*, *Lecture 2.*
[17] A. Smith, in J. M. Lothian (ed.), *op. cit.*, *Lecture 2.*
[18] http://www.electricscotland.com/history/other/smith_adam.htm

tion was already established – James Bowswell, Professor George Jardine, Henry Herbert, Lord Porchrester, and Hugh Blair were devoted participants in Smith's classes.

Undoubtedly, Smith's lectures on rhetoric and belle letters delivered in Edinburgh marked a clear transition from the earlier, well-established academic tradition of formal rhetoric. They were the first of the kind not only at the Edinburgh University but also in Great Britain giving the way to a more practical and creative attitude towards rhetoric. Smith's lectures showed that rhetoric was not merely a dull system of "silly books" but it was a critical study of literature and language aiming at the natural precision and simplicity of both spoken and written word. He departed from the stereotype of the lecturer who was merely a linguist and grammarian becoming also a literary critic, historian, stylist, writer, and orator. The direct consequence of the interest showed in the lectures was the creation of the Regius Chair of Rhetoric and Belle Lettres at the Edinburgh University which gave a continuity to Smith's ideas after his departure to Glasgow. The indirect and perhaps most important consequence of the lectures was the creation in Scotland of "an interested and informed audience for both what was creative and critical in literature".[19] Therefore, even though Dr Smith's lectures remain in the shadow of *The Wealth of Nations*, they have also done a great deal in shaping the Scottish intellectual mind.

[19] J. M. Lothian, *op. cit.*, p. XXIII.