Mohammed Abdellaoui
John D. Hey
Editors

SERIES C: Game Theory, Mathematical Programming and Operations Research 42

THEORY AND DECISION LIBRARY

Advances in Decision Making Under Risk and Uncertainty

Springer

THEORY AND DECISION LIBRARY

General Editors: W. Leinfellner (*Vienna*) and G. Eberlein (*Munich*)

Series A: Philosophy and Methodology of the Social Sciences

Series B: Mathematical and Statistical Methods

Series C: Game Theory, Mathematical Programming and Operations Research

Series D: System Theory, Knowledge Engineering and Problem Solving

---

SERIES C: GAME THEORY, MATHEMATICAL PROGRAMMING AND OPERATIONS RESEARCH

VOLUME 42

Editor-in-Chief: H. Peters (Maastricht University, The Netherlands); Honorary Editor: S. H. Tijs (Tilburg University, The Netherlands)
Editorial Board: E.E.C van Damme (Tilburg University, The Netherlands); H. Keiding (University of Copenhagen, Denmark); J.-F. Mertens (Université catholique de Louvain, Belgium); H. Moulin (Rice University, Houston, USA); Shigeo Muto (Tokyo University, Japan); T. Parthasarathy (Indian Statistical Institute, New Delhi, India); B. Peleg (Hebrew University, Jerusalem, Israel); T.E.S. Raghavan (University of Illinois at Chicago, USA); J. Rosenmüller (University of Bielefeld, Germany); A. Roth (Harvard University, USA); D. Schmeidler (Tel-Aviv University, Israel); R. Selten (University of Bonn, Germany); W. Thomson (University of Rochester, USA)

Scope: Particular attention is paid in this series to game theory and operations research, their formal aspects and their applications to economic, political and social sciences as well as to sociobiology. It will encourage high standards in the application of game-theoretical methods to individual and social decision making.

Mohammed Abdellaoui • John D. Hey
Editors

Advances in Decision Making Under Risk and Uncertainty

Springer

Editors

Prof. Mohammed Abdellaoui
Ecole des Hautes Etudes
Commerciales (HEC)
1, rue de la Libération
78351 Jouy-en-Josas
France
abdellaoui@hec.fr

Prof. Dr. John D. Hey
Department of Economics
University of York
Heslington, York Y01 5DD
United Kingdom
jdh1@york.ac.uk

ISBN: 978-3-540-68436-7 e-ISBN: 978-3-540-68437-4

Library of Congress Control Number: 2008929592

© 2008 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: WMX Design GmbH, Heidelberg

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

This volume contains a selection consisting of the best papers presented at the FUR XII conference, held at LUISS in Roma, Italy, in June 2006, organized by John Hey and Daniela Di Cagno. The objectives of the FUR (Foundations of Utility and Risk theory) conferences have always been to bring together leading academics from Economics, Psychology, Statistics, Operations Research, Finance, Applied Mathematics, and other disciplines, to address the issues of decision-making from a genuinely multi-disciplinary point of view. This twelfth conference in the series was no exception. The early FUR conferences – like FUR I (organized by Maurice Allais and Ole Hagen) and FUR III (organized by Bertrand Munier) – initiated the move away from the excessively rigid and descriptively-inadequate modelling of behaviour under risk and uncertainty that was in vogue in conventional economics at that time. More than twenty years later, things have changed fundamentally, and now innovations arising from the FUR conferences, and manifesting themselves in the new behavioural economics, are readily accepted by the profession. Working with new models of ambiguity, and bounded rationality, for example, behavioural decision making is no longer considered a sign of mere non-standard intellectual diversification. FUR XII was organised with this new spirit. In the sense that the behavioural concerns initiated by the first FUR conferences are now part of conventional economics, and the design and organisation of FUR XII reflects this integration, FUR XII represents a key turning point in the FUR conference series.

The 13 papers in this volume represent a sample of the best recent work in normative and descriptive modelling of behaviour under risk and uncertainty. We have divided the 13 papers into four broad parts (although there are obvious overlaps between the various parts): Uncertainty and information modelling; Risk modelling; Experimental individual decision making; and Experimental Interactive decision making.

Part I: Uncertainty and Information Modelling

There are four papers in this section. The one by Ghirardato et al. makes the fundamental claim that dynamic consistency – the fundamental property in dynamic choice models – is only compelling for choice situations in which acts are not affected by the possible presence of ambiguity. Their approach is based on one of the most general representations of preferences under uncertainty available up to now in the literature. Needless to say, such an approach opens new avenues of research on ambiguity. It also gives an edifying example of the maturity of research on decision making under uncertainty reached when FUR XII was organised.

Cohen et al. are also concerned with dynamic decision making under uncertainty but with exogenously given probabilities; they are interested in the role of risk perception. Their paper is another example of the use of insights from psychology and behavioural decision making in preference modelling.

Using a general framework of conditional preferences under uncertainty in the context of sequential equilibrium and rationalisability (building on earlier work by Asheim and Perea), Asheim shows that a conditional probability system (where each conditional belief is a subjective probability distribution) may lead to a refinement of a preference between two acts when new information – ruling out states at which the two acts coincide – becomes available.

Assuming that individual choice behaviour depends on more than the alternatives the decision maker is objectively facing, Stecher proposes an original axiomatic setup in which agents have preferences on their private subjective conceptions of possible alternatives. Given this axiomatic structure, the author provides conditions under which agents can communicate with others who do not necessarily perceive the world in the same way. The paper concludes that successful coordination needs the communication language between agents (for trade purposes) to be sufficiently vague. This is an important, if counter-intuitive, conclusion.

Part II: Risk Modelling

There are just three papers in this section. The first, one by Borgonovo and Peccati, works within the expected utility framework. They tackle sensitivity analysis as an integral part of any decision making process. Specifically, the authors answer two questions: the first concerning the response of decision making problems to small changes in the input (parameters); and the second relating to the problem of how the change is apportioned to input variations. The answers are important and interesting.

The second paper in the section is by Kaivanto and addresses the question of whether Cumulative Prospect Theory (CPT) resolves the famous St. Petersburg Paradox. Building on Rabin's "law of small numbers" (Rabin 2002), the author shows that the apparent failure of CPT popular parameterizations to resolve the paradox can be explained by the alternation bias inherent to the coin tossing process in the St. Petersburg gamble.

The final paper, one by Fabiyi, raises an interesting issue with respect to the form of the weighting function used in (Cumulative) Prospect Theory and in Rank Dependent Expected Utility function. Empirically it has often been observed to be S-shaped. Fabiyi provides a normative basis for this empirical finding.

Part III: Experimental Individual Decision Making

There are four papers in this section, illustrating the importance of experimental work and the amount of activity in this sector. The first is by Neugebauer who reports on an experiment in which the subject has to allocate his or her investment capital towards three assets. The experimental results confirm two main findings in behavioural decision making and behavioural finance – that is, first, that most subjects choose a dominated lottery when dominance is not transparent and, second, that subjects are loss-averse rather than variance-averse.

Carbone's contribution is concerned with the issue of dynamic inconsistencies and explores the possible influence of temptation as a reason for such inconsistencies. Motivated by the literature on hyperbolic discounting, she uses an innovative experimental design to investigate whether subjects are affected by temptation. The design involves an experiment with two treatments – one a 'spot market' and the other a 'forward market' – which should detect the existence of hyperbolicity. Interestingly, she finds little evidence of such behaviour.

Morone and Fiore report on an experiment in which the famous Monty Hall's three doors anomaly "should" go away. They deliberately adopt a design (Monty Hall's Three Doors for Dummies") which does not rely on subjects being able to do Bayesian updating. Nevertheless the anomaly does not go away – suggesting that the reasons for the anomaly are deeper and different than previously thought.

Giardini et al. argue, on the basis of two experimental studies using a 'visual motion discrimination task', that the desirability of an outcome may bias the amount of confidence people assign to the likelihood of that outcome. The originality of the authors' results lies in their observation that the correlation between reward and confidence was not linked to change in accuracy. In other words, subjects were not more accurate in responding to the stimulus; they were just more confident in their performance when facing a higher reward.

Part IV: Experimental Interactive Decision Making

The final section (on interactive experiments) contains three studies. That by Eichberger et al. extends the experimental study of ambiguity from individual decision making to interactive decision making (that is, to strategic games). The authors consider a non-standard situation in which players lack confidence in their equilibrium conjectures about opponents' play. They use "grannies, game theorists and fellow subjects" to introduce different levels of ambiguity in strategic games, and test comparative static propositions relating to changes in equilibrium with respect to changes in ambiguity.

Morone and Morone address the topic of guessing games with the objective of understanding whether people play in a rational or naïve way. They first develop a generalised theory of naïveté (that generalises the iterative naïve best replies strategy), and experimentally compare the iterative best replies strategy with the iterative elimination of dominated strategies for the generalised p-beauty contest.

Di Cagno and Sciubba explore network formation in a laboratory experiment. Instead of focusing on the traditional issue of convergence to a stable-network architecture, the authors use a network formation protocol suggesting that links are not unilateral, but have to be mutually agreed upon in order to form. The experimental results are analyzed from both 'macro' and 'micro' perspectives.

Taken together, the papers in this volume, a small subset of the papers presented at the 2006 FUR conference, show well what FUR is and what it does. We have already commented on the diversity of the papers in this volume, but the volume shows another facet of FUR – the desire and the ability to explore, both theoretically and empirically, new models of human behaviour. More importantly, as a study of the development of FUR over the years shows clearly, this volume manifests the clear and strong relationship between the theoretical and empirical developments: many of the empirical contributions would not have been possible without the earlier theoretical developments, and many of the theoretical papers are motivated by a desire to explain interesting phenomena thrown up by previous empirical papers. FUR demonstrates a strong commitment to interaction between theory and empirics. The editors of the present volume and the conference organizers are proud to contribute to keeping the FUR tradition alive.

Mohammed Abdellaoui
Jouy en Josas, April 2008
John D. Hey
York, April 2008

Contents

Contributors

Geir B. Asheim
Department of Economics, University of Oslo, P.O. Box 1095, Blindern, NO-0317 Oslo, Norway, g.b.asheim@econ.uio.no

Emanuele Borgonovo
IMQ, Bocconi University, Viale Isonzo 25, 20135, Milano, Italy, emanuele.borgonovo@unibocconi.it

Daniela Di Cagno
LUISS Guido Carli, Viale Pola 12, 00198 Roma, Italy, ddicagno@luiss.it

Enrica Carbone
Università di Bari, Facoltà di Economia, Via Camillo Rosalba 53, Italy, e.carbone@dse.uniba.it

Michèle Cohen
CES, Université de Paris I, 12-13 Boulevard de l'Hôpital, 75013 Paris, France, michele.cohen@univ-paris1.fr

Giorgio Coricelli
Institut des Sciences Cognitives, CNRS, 67, Boulevard Pinel, 69675, Bron, France, coricelli@isc.cnr.fr

Jürgen Eichberger
Wirtschaftstheorie I, Alfred-Weber-Institut f˙ur Wirtschaftswissenschaften, Grabengasse 14, 69117 Heidelberg, Germany, juergen.eichberger@awi.uni-heidelberg.de

Johanna Etner
CES, Université de Paris I, 12-13 Boulevard de l'Hôpital, 75013 Paris, France

Malcolm E. Fabiyi
University of Chicago Graduate School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA, mfabiyi@uchicago.edu

Annamaria Fiore
University of Bari, Via Camillo Rosalba 53, 70124 Bari, Italy, afiore@dse.uniba.it

Francesca Giardini
Institute of Cognitive Sciences and Technologies (ISTC), National Research Council, Via San Martino della Battaglia, 44, 00185, Rome, Italy, francesca.giardini@istc.cnr.it

Paolo Ghirardato
Collegio Carlo Alberto, Universit_a di Torino, Via Real Collegio, 30 10024 Moncalieri, Torino, Italy, paolo.ghirardato@unito.it

Meglena Jeleva
CES, Université de Paris I, 12-13 Boulevard de l'Hôpital, 75013 Paris, France, Meglena.Jeleva@univ-paris1.fr

Mateus Joffily
Institut des Sciences Cognitives, CNRS, 67, Boulevard Pinel, 69675, Bron, France, joffily@isc.cnr.fr

Kim Kaivanto
Department of Economics, Lancaster University, Management School, Lancaster LA1 4YX, UK, k.kaivanto@lancaster.ac.uk

David Kelsey
Department of Economics, School of Business and Economics, University of Exeter, Rennes Drive, Exeter, Devon EX4 4PU, UK, d.kelsey@exeter.ac.uk

Fabio Maccheroni
Dept. of Decision Sciences, Università Bocconi, Italy, fabio.maccheroni@unibocconi.it

Massimo Marinacci
Collegio Carlo Alberto, Universit_a di Torino Via Real Collegio, 30 10024 Moncalieri, Torino, Italy, massimo.marinacci@unito.it

Andrea Morone
Department of Economics, University of Bari, Via Camillo Rosalba 53, 70124 Bari, Italy, a.morone@gmail.com

Piergiuseppe Morone
Department of Economics, Mathematics and Statistics, University of Foggia, Largo Papa Giovanni Paolo II, 1 – 71100 - Foggia, Italy, p.morone@gmail.com

Tibor Neugebauer
University of Luxembourg, Luxembourg School of Finance, 4 rue Albert Borschette, 1246 Luxembourg, Luxembourg, Tibor.Neugebauer@uni.lu

Lorenzo Peccati
IMQ, Bocconi University, Viale Isonzo 25 20135, Milano, Italy, lorenzo.peccati@unibocconi.it

Angela Sirigu
Institut des Sciences Cognitives, CNRS, 67, Boulevard Pinel, 69675, Bron, France, sirigu@isc.cnr.fr

Burkhard C. Schipper
Department of Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA, bcschipper@ucdavis.edu

Jack Douglas Stecher
Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 Bergen, Norway, jack.stecher@nhh.no

Part I
Uncertainty and Information Modeling

Revealed Ambiguity and Its Consequences: Updating

P. Ghirardato(✉), F. Maccheroni, and M. Marinacci

Keywords: Ambiguity · Updating

1 Introduction

Dynamic consistency is a fundamental property in dynamic choice models. It requires that if a decision maker plans to take some action at some juncture in the future, he should consistently take that action when finding himself at that juncture, and vice versa if he plans to take a certain action at a certain juncture, he should take that plan in mind when deciding what to do now.

However compelling *prima facie*, it is well known in the literature that there are instances in which the presence of ambiguity might lead to behavior that reasonably violates dynamic consistency, as the next Ellsberg example shows.[1]

Example 1. Consider the classical "3-color" Ellsberg problem, in which an urn contains 90 balls, 30 of which are known to be red, while the remaining 60 are either blue or green. In period 0, the decision maker only has the information described above. Suppose that at the beginning of period 1 a ball is extracted from the urn, and the decision maker is then told whether the ball is blue or not. The decision maker has to choose between bets on the color of the drawn ball. Denoting by $[a,b,c]$ an act that pays a when a red ball is extracted, b when a green ball is extracted and c otherwise, let

$$f = [1,0,0]$$
$$g = [0,1,0]$$

P. Ghirardato
Collegio Carlo Alberto Via Real Collegio, 30 10024 Moncalieri (Torino), Italy
e-mail: paolo.ghirardato@unito.it

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.

© Springer-Verlag Berlin Heidelberg 2008

$$f' = [1,0,1]$$

$$g' = [0,1,1]$$

Suppose that in period 0, the decision maker, like most people in Ellsberg's experiment, displays the following preference pattern

$$g' \succ f' \succ f \succ g \tag{1}$$

(the middle preference being due to monotonicity). Letting $A = \{R, G\}$, it follows immediately from consequentialism that, conditionally on A^c,

$$f' \sim_{A^c} g'.$$

On the other hand, if the decision maker's conditional preferences satisfy dynamic consistency it must be the case that if he finds an act to be optimal conditionally on A and also conditionally on A^c in period 1, he must find the same act optimal in period 0. So, dynamic consistency implies that $g' \succ_A f'$ (as otherwise we should have $f' \succcurlyeq g'$). That is, a dynamically consistent and consequentialist decision maker who is told that a blue ball has not been extracted from the Ellsberg urn (i.e., is told A) must strictly prefer to bet on a green ball having been extracted.

Yet, it seems to us that a decision maker with the ambiguity averse preferences in (1) *might* still prefer to bet on a red ball being extracted, finding that event less ambiguous than the extraction of a green ball, and that constraining him to choose otherwise is imposing a strong constraint on the dynamics of his ambiguity attitude.

In view of this example, we claim that dynamic consistency is a compelling property only for comparisons of acts that are not affected by the possible presence of ambiguity. In other words, we think that rankings of acts unaffected by ambiguity should be dynamically consistent.

This is the starting point of this paper. We consider the preferences represented by

$$V(f) = a(f) \min_{P \in C} \int u(f(s))\, dP + (1 - a(f)) \max_{P \in C} \int u(f(s))\, dP, \tag{2}$$

where f is an act, a is a function over acts that describes the decision maker's attitudes toward ambiguity, and C is a set of priors that represents the ambiguity revealed by the decision maker's behavior. We provided an axiomatic foundation for such preferences in Ghirardato et al. (2004, henceforth GMM).[2] There, we also introduced a notion of unambiguous preference which is derived from the observable preference over acts. We argued that such derived unambiguous preference only ranks pairs of acts whose comparison is not affected by ambiguity. That is, unambiguous preference is a partial ordering, which is represented *à la* Bewley (2002) by the set of priors C (see (3) below).

Our main intuition is then naturally modelled by assuming that the derived unambiguous preference is dynamically consistent, while, in the presence of ambiguity, the primitive preference might well not be. This natural modelling idea leads to a simple and clean characterization of updating for the preferences we discuss in GMM. The main result of the present paper, Theorem 1, shows that the unambiguous preference is dynamically consistent if and only if all priors in C are updated according to Bayes' rule. This result thus characterizes prior by prior Bayesian updating, a natural updating rule for the preferences represented by (2).

We also consider a stronger dynamic consistency restriction on preferences, which can be loosely described as imposing dynamic consistency of the decision maker's "pessimistic self." We show that such restriction (unlike the one considered earlier) leads to imposing some structure on the decision maker's ex ante perception of ambiguity, which corresponds to the property that Epstein and Schneider (2003) have called *rectangularity*. This shows, *inter alia*, that rectangularity is not in general (i.e., for the preferences axiomatized in GMM) the characterization of dynamic consistency of the primitive preference relation, but of a different dynamic property which might even be logically unrelated to it.

We close by observing that we retain consequentialism of the primitive preference, another classic dynamic property that requires that preferences conditional on some event A only depend on the consequences inside A. This property has been weakened in Hanany and Klibanoff (2004), which also offers a survey of the literature on dynamic choice under ambiguity.

2 Preliminaries

2.1 Notation

Consider a set S of **states of the world**, an algebra Σ of subsets of S called **events**, and a set X of **consequences**. We denote by $\mathfrak{F}$ the set of all the **simple acts**: finite-valued Σ-measurable functions $f : S \to X$. Given any $x \in X$, we abuse notation by denoting $x \in \mathfrak{F}$ the constant act such that $x(s) = x$ for all $s \in S$, thus identifying X with the subset of the constant acts in $\mathfrak{F}$. Given $f, g \in \mathfrak{F}$ and $A \in \Sigma$, we denote by fAg the act in $\mathfrak{F}$ which yields $f(s)$ for $s \in A$ and $g(s)$ for $s \in A^c \equiv S \setminus A$. We model the DM's preferences on $\mathfrak{F}$ by a binary relation $\succcurlyeq$. As usual, $\succ$ and $\sim$ denote respectively the asymmetric and symmetric parts of $\succcurlyeq$.

We let $B_0(\Sigma)$ denote the set of all real-valued Σ-measurable simple functions, or equivalently the vector space generated by the indicator functions 1_A of the events $A \in \Sigma$. If $f \in \mathfrak{F}$ and $u : X \to \mathbb{R}$, we denote by $u(f)$ the element of $B_0(\Sigma)$ defined by $u(f)(s) = u(f(s))$ for all $s \in S$. A *probability charge* on (S, Σ) is function $P : \Sigma \to [0,1]$ that is normalized and (finitely) additive; i.e., $P(A \cup B) = P(A) + P(B)$ for any disjoint $A, B \in \Sigma$. Abusing our notation we sometimes use $P(\varphi)$ in place of $\int \varphi \, dP$, where $\varphi \in B_0(\Sigma)$.

Given a functional $I : B_0(\Sigma) \to \mathbb{R}$, we say that I is: **monotonic** if $I(\varphi) \geq I(\psi)$ for all $\varphi, \psi \in B_0(\Sigma)$ such that $\varphi(s) \geq \psi(s)$ for all $s \in S$; **constant additive** if $I(\varphi + \alpha) = I(\varphi) + \alpha$ for all $\varphi \in B_0(\Sigma)$ and $\alpha \in \mathbb{R}$; **positively homogeneous** if $I(\alpha\varphi) = \alpha I(\varphi)$ for all $\varphi \in B_0(\Sigma)$ and $\alpha \geq 0$; **constant linear** if it is constant additive and positively homogeneous.

Finally, as customary, given $f \in \mathfrak{F}$, we denote by $\Sigma(f)$ the algebra generated by f.

2.2 *Invariant Biseparable Preferences*

We next present the preference model used in the paper. We recall first the MEU model of Gilboa and Schmeidler (1989). In this model, a decision maker is represented by a utility function u and a set of probability charges $\mathcal{C}$, and she chooses according to the rule $\min_{P \in \mathcal{C}} \int u(\cdot)\, dP$. A generalization of this model is the so-called **α-maxmin** (α-MEU) model, in which the decision maker evaluates act $f \in \mathfrak{F}$ according to

$$\alpha \min_{P \in \mathcal{C}} \int_S u(f(s))\, dP(s) + (1 - \alpha) \max_{P \in \mathcal{C}} \int_S u(f(s))\, dP(s).$$

The α-MEU model is also a generalization – to an arbitrary set of priors, rather than the set of all possible priors on Σ – of Hurwicz's α-*pessimism* decision rule, which recommends evaluating an act by taking a convex combination (with weight α) of the utility of its worst possible result and of the utility of its best possible result. In collaboration with Arrow and Hurwicz (1972), Hurwicz later studied a generalization of his rule, which allows the "pessimism" weight α to vary according to the identity of the worst and best results that the act may yield.

As it turns out, there is a similar generalization of the α-MEU model allowing the weight α to depend on some features of the act f being evaluated. It is the model studied by GMM (see also Nehring (2001) and Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2003)), which relaxes Gilboa and Schmeidler's axiomatization of MEU by not imposing their "ambiguity aversion" axiom (and is constructed in a fully subjective setting). We present its functional characterization below, referring the reader to the cited Ghirardato et al. (2003, 2004) for the axiomatic foundation and further discussion. (The axioms are simply those of Gilboa and Schmeidler (1989) minus their "uncertainty aversion" axiom.)

Definition 1. A binary relation $\succcurlyeq$ on $\mathfrak{F}$ is called an **invariant biseparable preference** if there exist a unique monotonic and constant linear functional $I : B_0(\Sigma) \to \mathbb{R}$ and a nonconstant convex-ranged utility $u : X \to \mathbb{R}$, unique up to a positive affine transformation, such that $I(u(\cdot))$ represents $\succcurlyeq$; that is, for every $f, g \in \mathfrak{F}$,

$$f \succcurlyeq g \Leftrightarrow I(u(f)) \geq I(u(g)).$$

It is easy to see (see GMM, p. 157) that a functional $I : B_0(\Sigma) \to \mathbb{R}$ that satisfies monotonicity and constant linearity is also *Lipschitz continuous* of rank 1; i.e., $|I(\varphi) - I(\psi)| \leq \|\varphi - \psi\|$ for any $\varphi, \psi \in B_0(\Sigma)$.

In order to show how this model relates to the α-MEU model, we need to show how to derive a set of priors and consequently the decision maker's ambiguity attitude.

Suppose that act f is preferred to act g. If there is ambiguity about the state space, it is possible that such preference may not hold when we consider acts which average the payoffs of f and g with those of a common act h. Precisely, it is possible that a "mixed" act $g\,\lambda\,h$, which in each state s provides the average utility

$$u(g\,\lambda\,h)(s) = \lambda u(g(s)) + (1-\lambda)u(h(s)),$$

be preferred to a "mixed" act $f\,\lambda\,h$, which offers an analogous average of the payoffs of f and h. Such would be the case, for instance, if $g\,\lambda\,h$ has a utility profile which is almost independent of the realized state – while $f\,\lambda\,h$ does not – and the decision maker is pessimistic. On the other hand, there might be pairs of acts for which these "utility smoothing effects" are second-order. In such a case, we have "unambiguous preference." Precisely,

Definition 2. Let $f, g \in \mathfrak{F}$. Then, f is **unambiguously preferred** to g, denoted $f \succcurlyeq^* g$, if

$$f\,\lambda\,h \succcurlyeq g\,\lambda\,h$$

for all $\lambda \in (0,1]$ and all $h \in \mathcal{F}$.

Notice that in general $\succcurlyeq^*$ is a (possibly incomplete) coarsening of $\succcurlyeq$, while on the other hand for any $x, y \in X$, $x \succcurlyeq^* y$ if and only if $x \succcurlyeq y$.

In GMM we show that given an invariant biseparable preference there exists a unique nonempty, convex and (weak*) closed set $\mathcal{C}$ of probability charges that represents the unambiguous preference relation $\succcurlyeq^*$ in the following sense

$$f \succcurlyeq^* g \Longleftrightarrow \int_S u(f(s))\,dP(s) \geq \int_S u(g(s))\,dP(s) \qquad \text{for all } P \in \mathcal{C}. \tag{3}$$

That is, unambiguous preference corresponds to preference according to every one of the possible "probabilistic scenarios" included in $\mathcal{C}$. The set $\mathcal{C}$ therefore represents the ambiguity that is revealed by the decision maker's behavior.

Given the representation $\mathcal{C}$, the decision maker's index of ambiguity aversion a is then extracted from the functional I in the following natural way:

$$I(u(f)) = a(f) \min_{P \in \mathcal{C}} \int_S u(f(s))\,dP(s) + (1 - a(f)) \max_{P \in \mathcal{C}} \int_S u(f(s))\,dP(s).$$

The coefficient $a : \mathfrak{F} \to [0,1]$ is uniquely identified (GMM, Theorem 11) on the set of acts whose expectation is nonconstant over $\mathcal{C}$; i.e., those f for which it is *not* the case that

$$\int_S u(f(s))\,dP(s) = \int_S u(f(s))\,dQ(s) \text{ for every } P, Q \in \mathcal{C}. \tag{4}$$

Moreover, wherever uniquely defined, a also displays a significant regularity, as it turns out that $a(f) = a(g)$ whenever f and g "order" identically the possible scenarios in $\mathcal{C}$. Formally, for all $P, Q \in \mathcal{C}$,

$$\int_S u(f(s))\,dP(s) \geq \int_S u(f(s))\,dQ(s) \Longleftrightarrow \int_S u(g(s))\,dP(s) \geq \int_S u(g(s))\,dQ(s). \quad (5)$$

(See GMM, Proposition 10 and Lemma 8 respectively, for behavioral equivalents of the above conditions.) In words, the decision maker's degree of pessimism, though possibly variable, will not vary across acts which are symmetrically affected by ambiguity. Notice that in our environment the Arrow–Hurwicz rule corresponds to the case in which a decision maker's degree of pessimism only depends on the probabilities that maximize and minimize an act's evaluation. Thus, letting the degree of pessimism depend on all the ordering on $\mathcal{C}$ is a generalization of the Arrow–Hurwicz rule. Clearly, the SEU model corresponds to the special case in which $\mathcal{C}$ is a singleton. Thus, all SEU preferences whose utility is convex-ranged are invariant biseparable preferences. Less obviously, also CEU preferences with convex-ranged utility are invariant biseparable preferences. Hence, this model includes both α-MEU and CEU as special cases.

Unless otherwise noted, for the remainder of the paper preferences are always (but often tacitly) assumed to be invariant biseparable in the sense just described.

3 Some Derived Concepts

We introduce three notions which can be derived from the primitive preference relation via the unambiguous preference relation. Besides being intrinsically interesting, such notions prove useful in presenting the main ideas of the paper.

3.1 Mixture Certainty Equivalents

For any act $f \in \mathfrak{F}$, denote by $C^*(f)$ the set of the consequences that are "indifferent" to f in the following sense:

$$C^*(f) \equiv \{x \in X : \text{ for all } y \in X,\ y \succcurlyeq^* f \text{ implies } y \succcurlyeq^* x,\ f \succcurlyeq^* y \text{ implies } x \succcurlyeq^* y\}.$$

Intuitively, these are the constants that correspond to possible certainty equivalents of f. The set $C^*(f)$ can be characterized (GMM, Proposition 18) in terms of the set of expected utilities associated with $\mathcal{C}$:

Proposition 1. *For every* $f \in \mathfrak{F}$,

$$x \in C^*(f) \Longleftrightarrow \min_{P \in \mathcal{C}} P(u(f)) \leq u(x) \leq \max_{P \in \mathcal{C}} P(u(f)).$$

Moreover, $u(C^*(f)) = [\min_{P\in\mathcal{C}} P(u(f)), \max_{P\in\mathcal{C}} P(u(f))]$.

It follows immediately from the proposition that $x \in C^*(f)$ if and only if there is a $P \in \mathcal{C}$ such that $u(x) = P(u(f))$. That is, $u(C^*(f))$ is the range of the mapping that associates each prior $P \in \mathcal{C}$ with the expected utility $P(u(f))$.

There is another sense in which the elements of $C^*(f)$ are generalized certainty equivalents of f. Consider a consequence $x \in X$ that can be substituted to f as a "payoff" in a given mixture. That is, such that for some $\lambda \in (0,1]$ and $h \in \mathfrak{F}$,

$$x\lambda h \sim f\lambda h.$$

The following result shows that, while not all the elements of the set $C^*(f)$ can in general be expressed in this fashion, each of them is infinitesimally close (in terms of preference) to a consequence with this property.[3]

Proposition 2. *For every* $f \in \mathfrak{F}$, $C^*(f)$ *is the preference closure of the set*

$$\{x \in X : \exists\lambda \in (0,1], \exists h \in \mathfrak{F} \text{ such that } x\lambda h \sim f\lambda h\}.$$

In light of this result, we abuse terminology somewhat and call $x \in C^*(f)$ a **mixture certainty equivalent** of f, and $C^*(f)$ the **mixture certainty equivalents set** of f.

3.2 Lower and Upper Envelope Preferences

Given the unambiguous preference $\succcurlyeq^*$ induced by $\succcurlyeq$, we can also define the following two relations:

Definition 3. The **lower envelope preference** is the binary relation $\succcurlyeq^\downarrow$ on $\mathfrak{F}$ defined as follows: for all $f, g \in \mathfrak{F}$,

$$f \succcurlyeq^\downarrow g \Longleftrightarrow \{x \in X : f \succcurlyeq^* x\} \supseteq \{x \in X : g \succcurlyeq^* x\}.$$

The **upper envelope preference** is the binary relation $\succcurlyeq^\uparrow$ on $\mathfrak{F}$ defined as follows: for all $f, g \in \mathfrak{F}$,

$$f \succcurlyeq^\uparrow g \Longleftrightarrow \{x \in X : x \succcurlyeq^* f\} \subseteq \{x \in X : x \succcurlyeq^* g\}.$$

The relation $\succcurlyeq^\downarrow$ describes a "pessimistic" evaluation rule, while $\succcurlyeq^\uparrow$ an "optimistic" evaluation rule. To see this, notice that $\succcurlyeq^\downarrow$ ranks acts by the size of the set of consequences that are unambiguously worse than f. In fact, it ranks f exactly as the most valuable consequence that is unambiguously worse than f. The twin relation $\succcurlyeq^\uparrow$ does the opposite. We denote by $\succ^\downarrow$ and $\sim^\downarrow$ (resp. $\succ^\uparrow$ and $\sim^\uparrow$) the asymmetric and symmetric components of $\succcurlyeq^\downarrow$ (resp. $\succcurlyeq^\uparrow$) respectively.

This is further clarified by the following result, which shows that the envelope relations can be represented in terms of the set $\mathcal{C}$ derived in the previous section.

Proposition 3. *For every $f,g \in \mathfrak{F}$, the following statements are equivalent:*

(i) *$f \succcurlyeq^{\downarrow} g$ (resp. $f \succcurlyeq^{\uparrow} g$).*
(ii) $\min_{P\in\mathcal{C}} P(u(f)) \geq \min_{P\in\mathcal{C}} P(u(g))$ *(resp.* $\max_{P\in\mathcal{C}} P(u(f)) \geq \max_{P\in\mathcal{C}} P(u(g))$*).*

It follows from this result that $\succcurlyeq^{\downarrow}$ is a 1-MEU preference, in particular an invariant biseparable preference, and that $(\succcurlyeq^{\downarrow})^{*}$ is represented by $\mathcal{C}$. Moreover, while $\succcurlyeq$ and $\succcurlyeq^{\downarrow}$ always coincide on X, they coincide on $\mathfrak{F}$ if and only if $\succcurlyeq$ is 1-MEU, so that $\succcurlyeq$ and $\succcurlyeq^{\downarrow}$ will be in general distinct. Symmetric observations hold for $\succcurlyeq^{\uparrow}$.

The relations between $\succcurlyeq^{\downarrow}$, $\succcurlyeq^{\uparrow}$ and $\succcurlyeq$ can be better understood by recalling the relative ambiguity aversion ranking of Ghirardato and Marinacci (2002).

Proposition 4. *The preference relation $\succcurlyeq^{\downarrow}$ is more ambiguity averse than $\succcurlyeq$, which is in turn more ambiguity averse than $\succcurlyeq^{\uparrow}$.*

Therefore, the envelope relations can be interpreted as the "ambiguity averse side" and the "ambiguity loving side" of the DM. Indeed, $\succcurlyeq^{\downarrow}$ is ambiguity averse in the absolute sense of Ghirardato and Marinacci (2002), while $\succcurlyeq^{\uparrow}$ is ambiguity loving.

4 Revealed Ambiguity and Updating

Suppose that our DM has an information structure given by some subclass Π of Σ (say, a partition or a sub-algebra), and assume that we can observe our DM's ex ante preference on $\mathfrak{F}$, denoted interchangeably $\succcurlyeq$ or $\succcurlyeq_S$, and his preference on $\mathfrak{F}$ after having been informed that an event $A \in \Pi$ obtained, denoted $\succcurlyeq_A$. For each $A \in \Pi' \equiv \Pi \cup S$, the preference $\succcurlyeq_A$ is assumed to be invariant biseparable, and the utility representing $\succcurlyeq_A$ is denoted by u_A. Clearly, a conditional preference $\succcurlyeq_A$ also induces an unambiguous preference relation $\succcurlyeq_A^{*}$, as well as mixture certainty equivalents sets $C_A^{*}(\cdot)$ and a lower envelope preference relation $\succcurlyeq_A^{\downarrow}$. Because $\succcurlyeq_A$ is invariant biseparable, it is possible to represent $\succcurlyeq_A^{*}$ in the sense of (3) by a nonempty, weak* compact and convex set of probability measures $\mathcal{C}_A$.

We are interested in preferences conditional on events which are (ex ante) unambiguously non-null in the following sense:

Definition 4. We say that $A \in \Sigma$ is **unambiguously non-null** if $xAy \succ^{\downarrow} y$ for some (all) $x \succ y$.

That is, an event is unambiguously non-null if betting on A is unambiguously better than getting the loss payoff y for sure (notice that this is stronger than the definition of non-null event in Ghirardato and Marinacci (2001), which just requires that $xAy \succ y$). This property is equivalently restated in terms of the possible scenarios $\mathcal{C}$ as follows: $P(A) > 0$ for all $P \in \mathcal{C}$.

We next assume that conditional on being informed of A, the DM only cares about an act's results on A, a natural assumption that we call **consequentialism**: For every $A \in \Pi$, $f \sim_A fAg$ for every $f,g \in \mathfrak{F}$. Consequentialism extends immediately to the unambiguous and lower envelope preference relations, as the following result shows:

Lemma 1. *For every $A \in \Pi$, the following statements are equivalent*[4]*:*

(i) $f \sim_A f\ A g$ *for every* $f, g \in \mathfrak{F}$.
(ii) $f \sim_A^* f\ A g$ *for every* $f, g \in \mathfrak{F}$.
(iii) $f \sim_A^{\downarrow} f\ A g$ *for every* $f, g \in \mathfrak{F}$.

For the remainder of this section we tacitly assume that all the preferences $\succcurlyeq_A$ are invariant biseparable and consequentialist.

An important property linking ex ante and ex post preferences is **dynamic consistency**: For all $A \in \Pi$ and all $f, g \in \mathfrak{F}$,

$$f \succcurlyeq_A g \Longleftrightarrow f A g \succcurlyeq g. \tag{6}$$

This property imposes two requirements. The first says that the DM should consistently carry out plans made ex ante. The second says that information is valuable to the DM, in the sense that postponing her choice to after knowing whether an event obtained does not make her worse off (see Ghirardato (2002) for a more detailed discussion).

As announced in Sect. 1, we now inquire the effect of requiring dynamic consistency only in the absence of ambiguity; i.e., requiring (6) with $\succcurlyeq$ and $\succcurlyeq_A$ replaced by the unambiguous preference relations $\succcurlyeq^*$ and $\succcurlyeq_A^*$ respectively. We show that (for a preference satisfying consequentialism) this is tantamount to assuming that the DM updates all the priors in $\mathcal{C}$, a procedure that we call **generalized Bayesian updating**: For every $A \in \Pi$, the "updated" perception of ambiguity is equal to

$$\mathcal{C}|A \equiv \overline{co}^{w^*}\{P_A : P \in \mathcal{C}\},$$

where P_A denotes the posterior of P conditional on A, and $\overline{co}^{w^*}$ stands for the weak* closure of the convex hull.

Theorem 1. *Suppose that $A \in \Pi$ is unambiguously non-null. Then the following statements are equivalent:*

(i) *For every $f, g \in \mathfrak{F}$,*

$$f \succcurlyeq_A^* g \Longleftrightarrow P_A(u(f)) \geq P_A(u(g)) \textit{ for all } P \in \mathcal{C}. \tag{7}$$

Equivalently, $\mathcal{C}_A = \mathcal{C}|A$ and $u_A = u$.

(ii) *The relation $\succcurlyeq^*$ is dynamically consistent with respect to A. That is, for every $f, g \in \mathfrak{F}$:*

$$f \succcurlyeq_A^* g \Longleftrightarrow f A g \succcurlyeq^* g. \tag{8}$$

(iii) *For every $x, x' \in X$, $x \succcurlyeq x' \Rightarrow x \succcurlyeq_A x'$. For every $f \in \mathfrak{F}$ and $x \in X$:*

$$x \in C_A^*(f) \Longleftrightarrow x \in C^*(f A x). \tag{9}$$

(iv) *For every $f \in \mathfrak{F}$ and $x \in X$:*

$$f \succcurlyeq_A^{\downarrow} x \Longleftrightarrow f A x \succcurlyeq^{\downarrow} x. \tag{10}$$

Alongside the promised equivalence with dynamic consistency of unambiguous preference, this results presents two other characterizations of generalized Bayesian updating. They are inspired by a result of Pires (2002), who shows that when the primitive preference relations $\succsim_A$ are 1-MEU, generalized Bayesian updating is characterized by (a condition equivalent to)

$$f \succsim_A (\sim_A) x \Longleftrightarrow fAx \succsim (\sim) x \tag{11}$$

for all $f \in \mathfrak{F}$ and $x \in X$. Statement (iii) in the proposition departs from the indifference part of (11) and applies its logic to the "indifference" notion that is generated by the incomplete preference $\succsim^*$. Statement (iv) is a direct generalization of Pires's result to preferences that are not 1-MEU. Notice that (10) is equivalent to requiring that $f \succsim_A^* x$ if and only if $fAx \succsim^* x$, a weakening of (8) that under the assumptions of the proposition is equivalent to it.

It is straightforward to show that dynamic consistency of the primitives $\{\succsim_A\}_{A\in\Pi'}$ implies condition (ii). Thus, dynamic consistency of the primitives is a sufficient condition for generalized Bayesian updating. The following example reprises the Ellsberg discussion in Sect. 1 to show that it is not necessary.

Example 2. Consider the (CEU and) 1-MEU preference described by (linear utility and) the set $\mathcal{C} = \{P : P(R) = 1/3,\ P(G) \in [1/6, 1/2]\}$. It is clear that a decision maker with such $\mathcal{C}$ would display the preference pattern of (1). It follows from Theorem 1 that her preferences will satisfy consequentialism and unambiguous dynamic consistency if and only if conditionally on $A = \{R, G\}$ her updated set of priors is

$$\mathcal{C}_A = \{P : P(R) \in [2/5, 2/3]\}.$$

Assuming that the decision maker is also 1-MEU conditionally on A, this implies that in period 1 she will still prefer betting on a red ball over betting on a green ball. As discussed in Sect. 1, this cannot happen if the decision maker's conditional preferences satisfy dynamic consistency *tout court*; i.e., (6).

A different way of reinforcing the conditions of Theorem 1 is to consider imposing the full strength of dynamic consistency on the lower envelope preference relations, rather than the weaker form seen in (10). We next show that this leads to the characterization of the notion of rectangularity introduced by Epstein and Schneider (2003).

Suppose that the class Π forms a finite partition of S; i.e., $\Pi = \{A_1, \ldots, A_n\}$, with $A_i \cap A_j = \emptyset$ for every $i \neq j$ and $S = \cup_{i=1}^n A_i$. Given a set of probabilities $\mathcal{C}$ such that each A_i is unambiguously nonnull, we define

$$[\mathcal{C}] = \left\{ P : \exists Q, P_1, \ldots, P_n \in \mathcal{C} \text{ such that } \forall B \in \Sigma,\ P(B) = \sum_{i=1}^n P_i(B|A_i)\, Q(A_i) \right\}.$$

We say that $\mathcal{C}$ is **Π-rectangular** if $\mathcal{C} = [\mathcal{C}]$.[5] (We refer the reader to Epstein and Schneider (2003) for more discussion of this concept.)

Proposition 5. *Suppose that Π is a partition of S and that every $A \in \Pi$ is unambiguously non-null. Then the following statements are equivalent:*

(i) *$\mathcal{C}$ is Π-rectangular, and for every $A \in \Pi$, $u_A = u$ and $\mathcal{C}_A = \mathcal{C}|A$.*
(ii) *For every $f, g \in \mathfrak{F}$ and $A \in \Pi$:*

$$f \succsim_A^{\downarrow} g \Longleftrightarrow f A g \succsim^{\downarrow} g.$$

The rationale for this result is straightforward: Since the preference $\succsim^{\downarrow}$ is 1-MEU with set of priors $\mathcal{C}$, it follows from the analysis of Epstein and Schneider (2003) that $\mathcal{C}$ is rectangular and that for every $A \in \Pi$, $\mathcal{C}_A$ is obtained by generalized Bayesian updating. But the sets $\mathcal{C}_A$ are also those that represent the ambiguity perception of the primitive relations $\succsim_A$, as they represent the ambiguity perception of $\succsim_A^{\downarrow}$.

We have therefore shown that the characterization of rectangularity and generalized Bayesian updating of Epstein and Schneider can be extended to preferences which do not satisfy ambiguity hedging, having taken care to require dynamic consistency of the lower envelope (or equivalently of the upper envelope), rather than of the primitive, preference relations. The relations between dynamic consistency of the primitives $\{\succsim_A\}_{A \in \Pi'}$ and of the lower envelopes $\{\succsim_A^{\downarrow}\}_{A \in \Pi'}$ are not obvious and are an open research question.

Appendix

We begin with a preliminary remark and two pieces of notation, that are used throughout this appendix. First, notice that since $u(X)$ is convex, it is w.l.o.g. to assume that $u(X) \supseteq [-1, 1]$. Second, denote by $B_0(\Sigma, u(X))$ the set of the functions in $B_0(\Sigma)$ that map into $u(X)$. Finally, given a nonempty, convex and weak* compact set $\mathcal{C}$ of probability charges on (S, Σ), we denote for every $\varphi \in B_0(\Sigma)$,

$$\underline{\mathcal{C}}(\varphi) = \min_{P \in \mathcal{C}} P(\varphi), \qquad \overline{\mathcal{C}}(\varphi) = \max_{P \in \mathcal{C}} P(\varphi).$$

Proof of Proposition 2

Since the map from $B_0(\Sigma)$ to $\mathbb{R}$ defined by

$$\psi \mapsto I(u(f) + \psi) - I(\psi)$$

is continuous and $B_0(\Sigma)$ is connected, the set

$$\begin{aligned} J &= \{I(u(f) + \psi) - I(\psi) : \psi \in B_0(\Sigma)\} \\ &= \left\{ I\left(u(f) + \frac{1-\lambda}{\lambda} u(g)\right) - I\left(\frac{1-\lambda}{\lambda} u(g)\right) : g \in \mathfrak{F},\ \lambda \in (0,1] \right\} \end{aligned}$$

is connected. That is, it is an interval. From Lemma B.4 in GMM, it follows that

$$\overline{J} = [\underline{C}(u(f)), \overline{C}(u(f))].$$

Let

$$M(f) = \{x \in X : \exists \lambda \in (0,1], \exists h \in \mathfrak{F} \text{ such that } x\lambda h \sim f \lambda h\}.$$

We have $x \in M(f)$ iff

$$u(x) = I\left(u(f) + \frac{1-\lambda}{\lambda}u(h)\right) - I\left(\frac{1-\lambda}{\lambda}u(h)\right)$$

iff $x \in u^{-1}(J)$. Hence, $u(M(f)) \subseteq J$. Conversely, if $t \in J$, $t \in [\underline{C}(u(f)), \overline{C}(u(f))]$ and there exists $x \in X$ such that $u(x) = t$. Clearly, $x \in M(f)$, whence $u(M(f)) = J$. We conclude observing that

$$C^*(f) = u^{-1}\left([\min_{P \in C} P(u(f)), \max_{P \in C} P(u(f))]\right) = u^{-1}\left(\overline{J}\right) = u^{-1}\left(\overline{u(M(f))}\right).$$

Proof of Proposition 3

To prove the statement for $\succcurlyeq^{\downarrow}$ (that for $\succcurlyeq^{\uparrow}$ is proved analogously), we only need to show that

$$f \succcurlyeq^{\downarrow} g \Longleftrightarrow \underline{C}(u(f)) \geq \underline{C}(u(g)).$$

Applying the definition of $\succcurlyeq^{\downarrow}$ and the representation of (3), we have that $f \succcurlyeq^{\downarrow} g$ iff for every $x \in X$,

$$P(u(g)) \geq u(x) \text{ for all } P \in C \Rightarrow P(u(f)) \geq u(x) \text{ for all } P \in C.$$

That is, iff for every $x \in X$,

$$\underline{C}(u(g)) \geq u(x) \Rightarrow \underline{C}(u(f)) \geq u(x).$$

This is equivalent to $\underline{C}(u(f)) \geq \underline{C}(u(g))$, concluding the proof.

Proof of Proposition 4

We have proved in Proposition 3 that $\succcurlyeq^{\downarrow}$ is represented by the functional $\underline{C}(u(\cdot))$, and $\succcurlyeq^{\uparrow}$ by $\overline{C}(u(\cdot))$. Consider $\succcurlyeq$ and $\succcurlyeq^{\downarrow}$, and notice that (GMM, Proposition 7) for any $f \in \mathfrak{F}$, $\underline{C}(u(f)) \leq I(u(f)) \leq \overline{C}(u(f))$. It is clear that $\underline{C}(u(f)) \leq I(u(f))$ is tantamount to saying that for every $x \in X$,

$$x \succcurlyeq f \Rightarrow x \succcurlyeq^{\downarrow} f.$$

The argument for $\succcurlyeq$ and $\succcurlyeq^{\uparrow}$ is analogous.

Proof of Lemma 1

$(i) \Leftrightarrow (ii)$: Assume that, for every $A \in \Pi$, $f \sim_A fAg$ for every $f,g \in \mathfrak{F}$, hence, for every $h \in \mathfrak{F}$ $[\lambda f + (1-\lambda)h] \sim_A [\lambda f + (1-\lambda)h]A[\lambda g + (1-\lambda)h]$, that is $\lambda f + (1-\lambda)h \sim_A \lambda fAg + (1-\lambda)h$, thus $f \sim_A^* fAg$. Conversely, if $f \sim_A^* fAg$ for every $f,g \in \mathfrak{F}$, then in particular $f \sim_A fAg$ for every $f,g \in \mathfrak{F}$.

$(ii) \Leftrightarrow (iii)$: By (3), $f \sim_A^* fAg$ iff $P(u_A(f)) = P(u_A(fAg))$ for all $P \in \mathcal{C}_A$. It immediately follows that $\underline{\mathcal{C}_A}(u_A(f)) = \underline{\mathcal{C}_A}(u_A(fAg))$. By Proposition 3, this is equivalent to $f \sim_A^{\downarrow} fAg$.

Conversely, suppose that $f \sim_A^{\downarrow} fAg$ for every $f,g \in \mathfrak{F}$. Consider $x \succ_A y$. Since $x \sim_A^{\downarrow} xAy$, it follows from Proposition 3 that

$$\begin{aligned} u_A(x) &= \min_{P \in \mathcal{C}_A} [u_A(x)P(A) + u_A(y)(1-P(A))] \\ &= u_A(x) \min_{P \in \mathcal{C}_A} P(A) + u_A(y)(1 - \min_{P \in \mathcal{C}_A} P(A)). \end{aligned}$$

Since $u_A(x) > u_A(Y)$, this implies that $\min_{P \in \mathcal{C}_A} P(A) = 1$, or equivalently, that $P(A) = 1$ for all $P \in \mathcal{C}_A$. It follows that $P(u_A(f)) = P(u_A(fAg))$ for all $P \in \mathcal{C}_A$, which is equivalent to $f \sim_A^* fAg$.

Proof of Theorem 1

First, we observe that the fact that (7) implies $\mathcal{C}_A = \mathcal{C}|A$ is a consequence of Proposition A.1 in GMM. That it implies $u_A = u$ is seen by taking $f = x$ and $g = x'$ to show that $x \succsim_A x' \Leftrightarrow x \succsim x'$. The converse is trivial.

$(i) \Leftrightarrow (ii)$: $fAg \succsim^* g$ for all $P \in \mathcal{C}$ iff $\int_A u(f)\,dP + \int_{A^c} u(g)\,dP \geq \int_A u(g)\,dP + \int_{A^c} u(g)\,dP$ for all $P \in \mathcal{C}$ iff $\int_A u(f)\,dP \geq \int_A u(g)\,dP$ for all $P \in \mathcal{C}$ iff $P_A(u(f)) \geq P_A(u(g))$ for all $P \in \mathcal{C}$.

$(i) \Rightarrow (iii)$: Suppose that $u = u_A$. We first observe that it follows from Proposition 1 that for every $f \in \mathfrak{F}$ and $x \in X$, with obvious notation,

$$x \in C_A^*(f) \Longleftrightarrow \underline{\mathcal{C}_A}(u(f)) \leq u(x) \leq \overline{\mathcal{C}_A}(u(f)). \tag{12}$$

Next, we prove that for every $f \in \mathfrak{F}$ and $x \in X$, again with obvious notation,

$$x \in C^*(fAx) \Longleftrightarrow \underline{\mathcal{C}|A}(u(f)) \leq u(x) \leq \overline{\mathcal{C}|A}(u(f)). \tag{13}$$

To see this, apply again Proposition 1 to find

$$x \in C^*(fAx) \Longleftrightarrow \underline{\mathcal{C}}(u(fAx)) \leq u(x) \leq \overline{\mathcal{C}}(u(fAx)).$$

That is, $x \in C^*(fAx)$ iff both

$$\min_{P \in \mathcal{C}} \int_S u(fAx)\,dP \leq u(x) \tag{14}$$

and

$$u(x) \leq \max_{P \in \mathcal{C}} \int_S u(fAx)\, dP. \tag{15}$$

Denote resp. $\underline{P}$ and $\overline{P}$ the probabilities in $\mathcal{C}$ that attain the extrema in (14) and (15). Then we can rewrite (14) as follows:

$$u(x) \geq \frac{1}{\underline{P}(A)} \int_A u(f)\, d\underline{P},$$

which is equivalent to saying that

$$u(x) \geq \min_{P \in \mathcal{C}} \int_S u(f)\, dP_A = \underline{\mathcal{C}|A}(u(f)).$$

Analogously, (15) can be rewritten as

$$u(x) \leq \frac{1}{\overline{P}(A)} \int_A u(f)\, d\overline{P},$$

which is equivalent to

$$u(x) \leq \max_{P \in \mathcal{C}} \int_S u(f)\, dP_A = \overline{\mathcal{C}|A}(u(f)).$$

This ends the proof of (13).

To prove (i), notice that $x \succcurlyeq x' \Rightarrow x \succcurlyeq_A x'$ obviously follows from the assumption $u = u_A$, and that, given (12) and (13), (9) follows immediately from the assumption $\mathcal{C}|A = \mathcal{C}_A$.

$(iii) \Rightarrow (i)$: First, observe that the assumption that $x \succcurlyeq x' \Rightarrow x \succcurlyeq_A x'$ implies $u = u_A$ by Corollary B.3 in GMM. Hence, it follows from (12) and (13) above that (9) is equivalent to

$$\underline{\mathcal{C}_A}(u(f)) \leq u(x) \leq \overline{\mathcal{C}_A}(u(f)) \Longleftrightarrow \underline{\mathcal{C}|A}(u(f)) \leq u(x) \leq \overline{\mathcal{C}|A}(u(f)).$$

In particular, this implies that for every $\varphi \in B_0(\Sigma, u(X))$,

$$\min_{P \in \mathcal{C}|A} P(\varphi) = \min_{Q \in \mathcal{C}_A} Q(\varphi) \tag{16}$$

The result that $\mathcal{C}|A = \mathcal{C}_A$ now follows from two applications of Proposition A.1 in GMM.

$(i) \Rightarrow (iv)$: By Proposition 3 and the assumption that $u_A = u$, for every $f \in \mathfrak{F}$ and $x \in X$ we have that $f \succcurlyeq_A^{\downarrow} x$ iff $Q(u(f)) \geq u(x)$ for all $Q \in \mathcal{C}_A$. Next, we show that

$$fAx \succcurlyeq^{\downarrow} x \Longleftrightarrow P(u(f)) \geq u(x) \quad \text{for all } P \in \mathcal{C}|A, \tag{17}$$

so that the result follows from the assumption that $\mathcal{C}|A = \mathcal{C}_A$.

To see why (17) holds, notice that by Proposition 3, $fAx \succcurlyeq^{\downarrow} x$ iff $P(u(fAx)) \geq u(x)$ for all $P \in \mathcal{C}$. Equivalently, for every $P \in \mathcal{C}$,

$$\int_A u(f)\, dP + (1 - P(A))\, u(x) \geq u(x),$$

which holds iff $P_A(u(f)) \geq u(x)$ (recall that $P(A) > 0$ for all $P \in \mathcal{C}$). In turn, the latter is equivalent to saying that $P(u(f)) \geq u(x)$ for every $P \in \mathcal{C}|A$.

$(iv) \Rightarrow (i)$: We first show (mimicking an argument of Siniscalchi (2001)) that (10) implies that $u_A = u$. To see this, notice that we have $u_A(x) \geq u_A(x')$ iff $x \succcurlyeq_A x'$ iff $xAx' \succcurlyeq^{\downarrow} x'$ iff

$$\min_{P \in \mathcal{C}} \left[u(x)P(A) + u(x')(1 - P(A)) \right] \geq u(x').$$

Using the assumption that $\min_{P \in \mathcal{C}} P(A) > 0$, the latter is equivalent to $u(x) \geq u(x')$, proving that $u_A = u$.

We now show that $\mathcal{C}|A = \mathcal{C}_A$ by showing that (16) holds for every $\varphi \in B_0(\Sigma, u(X))$, so that the result follows again from Proposition A.1 in GMM. As argued above, (10) holds for f and x iff

$$P(u(f)) \geq u(x) \quad \text{for all } P \in \mathcal{C}|A \Longleftrightarrow Q(u(f)) \geq u(x) \quad \text{for all } Q \in \mathcal{C}_A.$$

Fix $\varphi \in B_0(\Sigma, u(X))$ and suppose that, in violation of (16), $\alpha \equiv \min_{P \in \mathcal{C}|A} P(\varphi) > \min_{Q \in \mathcal{C}_A} Q(\varphi) \equiv \beta$. Then there exists $\gamma \in (\beta, \alpha)$. Let x denote the consequence such that $u(x) = \gamma$. By the assumption we just made, we have that (if $f \in \mathfrak{F}$ is such that $u(f) = \varphi$),

$$\min_{P \in \mathcal{C}|A} P(u(f)) > u(x) \quad \text{and} \quad u(x) > \min_{Q \in \mathcal{C}_A} Q(u(f)),$$

which, as proved above (the proof for strict preference works *mutatis mutandis* as that for weak preference, recalling that $\mathcal{C}|A$ is weak* compact), is equivalent to $fAx \succ^{\downarrow} x$ and $f \prec_A^{\downarrow} x$, a contradiction. Suppose instead that $\alpha < \beta$ and let $\gamma \in (\alpha, \beta)$. In this case we obtain

$$\min_{P \in \mathcal{C}|A} P(u(f)) < u(x) \quad \text{and} \quad u(x) < \min_{Q \in \mathcal{C}_A} Q(u(f)),$$

which is equivalent to $f \succ_A^{\downarrow} x$ and $x \succ^{\downarrow} fAx$ (to see the latter, let $P^* \in \mathcal{C}$ be the probability whose posterior minimizes the left-hand inequality; it follows that $\min_{P \in \mathcal{C}} P(u(fAx)) \leq P^*(u(fAx)) < u(x)$), again a contradiction. This concludes the proof of (16).

Acknowledgements Most of the results in this paper were earlier circulated as part of a working paper titled "Ambiguity from the Differential Viewpoint" (Caltech Social Sciences Working Paper Number 1130, April 2002). We are grateful to Marciano Siniscalchi for many conversations on the topic of this paper.

Notes

[1]We owe this example to Denis Bouyssou, who showed it to us at the RUD 1997 conference in Chantilly.

[2]The multiple priors model of Gilboa and Schmeidler (1989) corresponds to the special case in which $a(f) = 1$ for all acts f. The Choquet expected utility model of Schmeidler (1989) is also seen to be a special case.

[3]In the statement by "preference closure" of a subset $Y \subseteq X$, we mean $u^{-1}\left(\overline{u(Y)}\right)$.

[4]In this and the remaining results of this section, we omit the equivalent statements involving the upper envelope preference relation.

[5]We owe this presentation of rectangularity to Marciano Siniscalchi.

References

Arrow, K. J., & Hurwicz, L. (1972). An optimality criterion for decision making under ignorance. In C. Carter & J. Ford (Eds.), *Uncertainty and expectations in economics*. Oxford: Basil Blackwell.

Bewley, T. (2002, November). Knightian decision theory: Part I. *Decisions in Economics and Finance, 25*(2), 79–110. ((first version 1986))

Epstein, L. G., & Schneider, M. (2003, August). Recursive multiple-priors. *Journal of Economic Theory, 113*, 1–31.

Ghirardato, P. (2002, March). Revisiting Savage in a conditional world. *Economic Theory, 20*, 83–92.

Ghirardato, P., Maccheroni, F., & Marinacci, M. (2004, October). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory, 118*(2), 133–173.

Ghirardato, P., Maccheroni, F., Marinacci, M., & Siniscalchi, M. (2003, November). A subjective spin on roulette wheels. *Econometrica, 17*(6), 1897–1908.

Ghirardato, P., & Marinacci, M. (2001, November). Risk, ambiguity, and the separation of utility and beliefs. *Mathematics of Operations Research, 26*(4), 864–890.

Ghirardato, P., & Marinacci, M. (2002, February). Ambiguity made precise: A comparative foundation. *Journal of Economic Theory, 102*, 251–289.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics, 18*, 141–153.

Hanany, E., & Klibanoff, P. (2004, June). *Updating preferences with multiple priors multiple priors*. (Mimeo, Northwestern University)

Nehring, K. (2001, November). *Ambiguity in the context of probabilistic beliefs*. (Mimeo, UC Davis)

Pires, C. P. (2002). A rule for updating ambiguous beliefs. *Theory and Decision, 53*, 137–152.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica, 57*, 571–587.

Siniscalchi, M. (2001, September). *Bayesian updating for general maxmin expected utility preferences*. (Mimeo, Princeton University)

Dynamic Decision Making When Risk Perception Depends on Past Experience

M. Cohen(✉), J. Etner, and M. Jeleva

Keywords: Dynamic decision making · Past experience · Rank dependent utility model · Recursive model · Risk perception

1 Introduction

Decision theory under risk had for a long time focused mainly on the impact of different risk and wealth perceptions on the agents' optimal decisions. In these classical studies, risk perceptions, as well as utility functions, depend only on the considered decision characteristics (pairs probabilities-outcomes) and thus cannot be influenced by outside factors. However, some psychological studies (see Slovic (2000)) point out the fact that risk perception may be strongly influenced by the context in which the individuals are when they take their decisions. Context can take different forms: (a) it can correspond to past experience (relevant, for instance, in insurance decisions, as noticed in Kunreuther (1996) and Browne and Hoyt (2000) and in stock markets behavior as noticed by Hirshleifer and Shumway (2003)), (b) it can also correspond to anticipatory feelings about some future states (Caplin and Leahy 2001) (c) it can be related to the decision outcomes presentation (leading then to the framing effect pointed out by Tversky and Kahneman (1986)).

In this paper, we focus on the context generated by past experience, corresponding to a sequence of events occurring before the moment of the decision. This past experience can concern different events: (a) past realizations of the decision-relevant events (as accidents when an insurance decision is considered) or (b) past realizations of other events (such as weather conditions when a stock market behavior is considered).

M. Cohen
PSE and CES University Paris 1, MSE, 106-112 Bld de I'Hôpital, 75013 Paris, France
e-mail: michele.cohen@univ-paris1.fr

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.
© Springer-Verlag Berlin Heidelberg 2008

The influence of past experience on decisions appears in particular on insurance markets for catastrophic risk. It appears that in California, before the earthquake in 1989, 34% of the individuals consider that insurance against earthquake is useless; after the earthquake, they are only 5% to have this opinion. Moreover, the earthquake occurrence increases insurance demand: 11% of the non insured individuals subscribed an insurance contract (Kunreuther 1996). These results are confirmed by an empirical study from Browne and Hoyt (2000) that reveals a strong positive correlation between the number of flood insurance contracts subscribed in a year in a given State of the US and losses due to flood in the same State the previous year. A relation between past experience and insurance demand appears also in experimental studies when individuals are well informed about the probability of loss realization and about the independence of losses in successive periods (McClelland, Schulze, & Coursey 1993; Ganderton, Brookshire, McKee, Stewart, & Thurston 2000; Papon 2004). However, the results are less clear-cut: if the existence of a strong correlation between past damages and insurance demand is well established, its sign is less clear. Indeed, two opposite effects can be identified corresponding to availability bias and gambler fallacy in the sense of Tversky and Kahneman (1973). The availability bias corresponds to an overestimation of the probability of an event that recently occurred and implies an increase in insurance demand after a natural disaster, this demand being low after a long period without a catastrophe. The opposite occurs with the gambler's fallacy effect: individuals underestimate the probability of repetition of the event that they just observed and thus buy less insurance after a catastrophe.

When events are independent over time, behaviors that we have just described cannot be explained in the standard expected utility model. Indeed, in this model, past experienced losses lead only to a wealth decrease and not to a probability assessment modification.

Relaxing the axiom of context independence of preferences under uncertainty can allow the rationalization of some decisions considered as inconsistent with respect to the existing criteria because reflecting unstable preferences as for instance the modification of insurance demand against catastrophic risk after the occurrence of a catastrophe.

The aim of the paper is to propose a preferences representation model under risk where risk perception can be past experience dependent. A first step consists in considering a one period decision problem where individual preferences are no more defined only on decisions but on pairs (decision, past experience). The obtained criterion is used in the construction of a dynamic choice model under risk.

The underlying model of decision making under risk that is used here is the RDU (Rank Dependent Utility), proposed by Quiggin (1982) and Yaari (1987). This model has the advantage to allow a non linear treatment of probabilities, in addition to a non linear treatment of outcomes. When one period decisions are considered, we adapt the RDU axiomatic system of Chateauneuf (1999) to represent preferences on pairs (decision, past experience).

To better capture the long term impact of past experience on decisions, after the preferences representation on a point of time, we model intertemporal decisions.

RDU model can generate dynamic inconsistency. To rule out this problem, we use the recursive model of Kreps and Porteus (1978). In the latter model, risk aversion is characterized by a standard utility function and the agents' past experience is summarized by a sequence of monetary payoffs, resulting from the past decisions and the lottery realizations. In the present paper are introduced additional aspects: (a) probabilities treatment is non linear; (b) past experience does not reduce anymore to the only payoffs, but is characterized by a more general sequence of events, related or not to the decision relevant events. To achieve this preferences representation, we assume that preferences at a point of time are represented by the "past experience dependent" RDU previously axiomatized and modify the dynamic consistency axiom of KP in order to apply to states and not to payments.

The paper starts with the "past experience dependent" preferences representation at a point of time. We propose an axiomatic foundation for "past experience dependent" rank dependent utility under risk. In Section 3, we consider a dynamic choice problem and prove a representation theorem. Section 4 contains an illustrative example.

2 Behavior at a Point of Time

In this section, we consider a static problem. We propose an axiomatic representation of preferences by a rank dependent expected utility which takes into account the agent's past experience.

2.1 Notations and Definitions

Decision problem is characterized by a set of risky perspectives in which an agent has to make his choice and by a set of states that characterize the agent's past experience.

Let $\mathcal{Z}$ denote a set of outcomes. We assume that $\mathcal{Z}$ is a non empty connected compact and metric space and $\mathcal{L}$ is the set of lotteries over $\mathcal{Z}$.

$\mathcal{S}$ is a set of realized states, assumed nonempty, compact, connected separable topological space. An element of $\mathcal{S} \times \mathcal{L}$ will be called a "past experience dependent lottery".

$\succsim$ is a binary relation on $\mathcal{S} \times \mathcal{L}$ which denotes the preference relation of a decision maker. We denote by $\succ$ the strict preference and by $\sim$ the indifference.

Axiom 1 $\succsim$ *is a weak order on* $\mathcal{S} \times \mathcal{L}$.

The preferences representation of $\succsim$ on $\mathcal{S} \times \mathcal{L}$ will be built in two steps. We start with the preference relation on $s \times \mathcal{L}$ and its representation by a Rank Dependant Utility (RDU) model. Then, we give some additional assumptions to achieve a RDU preferences representation on $\mathcal{S} \times \mathcal{L}$.

2.2 Preferences on $s \times \mathcal{L}$ and RDU

In this section we consider the restrictions of $\succsim$ to $s \times \mathcal{L}$ that we denote by $\succsim_s$. For a given state s, we face a standard decision problem under risk.

The RDU representation of the preference relation $\succsim_s$ is obtained by the following axioms, proposed in Chateauneuf (1999).

From Axiom 1, it follows directly that $\succsim_s$ is a weak order on $s \times \mathcal{L}$.

Axiom 2 *(Continuity) For a given $s \in \mathcal{S}$, let $P_n = (s, L_n), P = (s, L), Q = (s, L') \in s \times \mathcal{L}$, with P_n weakly converging to P, then $\forall n, P_n \succsim_s Q \Rightarrow P \succsim_s Q$ and $\forall n, P_n \precsim_s Q \Rightarrow P \precsim_s Q$.*

For a given s, it is possible to completely order the space $\mathcal{Z}$. The definition of the first order stochastic dominance (FSD) becomes $L\ FSD\ L'$ if and only if $P_L(z \in \mathcal{Z},\ (s,z) \succ_s (s,x)) \geq P_{L'}(z \in \mathcal{Z},\ (s,z) \succ_s (s,x))\ \forall x \in \mathcal{Z}$.

The next axiom guarantees that $\succsim_s$ preserves first order stochastic dominance.

Axiom 3 *For any $L, L' \in \mathcal{L}$ such that $L\ FSD\ L'$, $(s,L) \succsim_s (s,L')$.*

Axiom 4 (Comonotonic Sure-Thing Principle) *For any $s \in \mathcal{S}$, let lotteries $P = \left(\left(s, z^i\right) p_i\right), Q = \left(\left(s, y^i\right) p_i\right)$ be such $(s, z^{i_0}) \sim_s \left(s, y^{i_0}\right)$, then $P \succsim_s Q$ implies $P' \succsim_s Q'$, for lotteries P', Q' obtained from lotteries P and Q by merely replacing the i_0^{th} common pair (s, z^{i_0}), by a common pair (s, x^{i_0}) again in i_0^{th} rank both in P' and Q'.*

Axiom 5 (Comonotonic Mixture Independance Axiom) *For any $s \in \mathcal{S}$, and for any lotteries $P = \left(\left(s, z^i\right) p_i\right)$ and $Q = \left(\left(s, y^i\right) q_i\right)\right)$,*

For any $p \in [0,1]$, for any $a, b, c, d \in \mathcal{Z}$

- $P_1 = (1-p)\left(s, z^{\min}\right) + p(s,a) \sim_s Q_1 = (1-p)\left(s, y^{\min}\right) + p(s,b)$

  and $P_2 = (1-p)\left(s, z^{\min}\right) + p(s,c) \sim_s Q_2 = (1-p)\left(s, y^{\min}\right) + p(s',d)$
  imply $\forall \alpha \in [0,1], \alpha P_1 + (1-\alpha) P_2 \sim_s \alpha Q_1 + (1-\alpha) Q_2$

- $P_1 = (1-p)(s, z^{\max}) + p(s,a) \sim_s Q_1 = (1-p)(s, y^{\max}) + p(s',b)$

  and $P_2 = (1-p)(s, z^{\max}) + p(s,c) \sim_s Q_2 = (1-p)(s, y^{\max}) + p(s',d)$
  imply $\forall \alpha \in [0,1], \alpha P_1 + (1-\alpha) P_2 \sim_s \alpha Q_1 + (1-\alpha) Q_2$

Theorem 1. *Let the preference relation $\succsim_s$ on $s \times \mathcal{L}$ satisfy Axioms 1–5, then there exist an increasing function $\varphi_s : [0,1] \to [0,1]$, with $\varphi_s(0) = 0$, $\varphi_s(1) = 1$ and a utility function, $v_s : \mathcal{Z} \to \mathbf{R}$, which is increasing, continuous, and unique up to an affine transformation such that:*

$\forall L, L' \in \mathcal{L}$, $(s,L) \succsim_s (s,L')$ *iff* $V_s(L) \geq V_s(L')$ *with*

$$V_s(L) = \sum_{i=1}^{n} \left(\varphi_s\left(\sum_{j=i}^{n} p_j\right) - \varphi_s\left(\sum_{j=i+1}^{n} p_j\right) \right) \times v_s(z^i)$$

Proof. Chateauneuf (1999). □

We now consider the general preferences on $\mathcal{S} \times \mathcal{L}$ comparing lotteries in different contexts. Let us notice that the preferences on $\mathcal{S} \times \mathcal{L}$ induce a preference on $\mathcal{S} \times \mathcal{Z}$ but in no case a preference on $\mathcal{S}$ alone.

At this stage, the payoffs evaluation depends not only on z but also on s. The objective of this paper is to emphasize the link between risk perception and individual context. In order to isolate this feature, we assume that only risk perceptions depend on s. This assumption needs more discussion, mainly with respect to the state-dependant model of Karni (1985). The main feature of Karni's model is that the evaluation of a given amount of money may strongly depend on the state in which the individual is when receiving this amount. Here, the states we consider are not of the same type: they are already realized (past) states, and not future ones. It seems then more realistic to assume that they are more likely to influence risk perception than future monetary evaluations.

The following axiom guarantees that the payoffs evaluations do not depend on past experience.

Axiom 6 *For any $s, s' \in \mathcal{S}$ and any $z \in \mathcal{Z}$, $(s,z) \sim (s',z)$.*

Let us notice that Axioms 3 and 6 induce the existence of a preference relation on $\mathcal{Z}$ independent of $\mathcal{S}$. To simplify notations, we can then write $z \geq z'$ instead of $(s,z) \succeq (s,z')$ for all $s \in \mathcal{S}$.

The following preferences representation theorem can then be formulated.

Theorem 2. *Under Axioms 1–6, the weak order $\succsim$ on $\mathcal{S} \times \mathcal{L}$ is representable by a function $V : \mathcal{S} \times \mathcal{L} \to \mathbf{R}$. For any $s, s' \in \mathcal{S}$ and any $L, L' \in \mathcal{L}$, $(s,L) \succsim (s',L') \Leftrightarrow V(s,L) \geq V(s',L')$*

where $V(s,L) = \sum\limits_{i=1}^{n} \left(\varphi_s(\sum\limits_{j=i}^{n} p_j) - \varphi_s(\sum\limits_{j=i+1}^{n} p_j) \right) v(z^i).$

Proof. The generalization of the preferences representation of the restrictions $\succsim_s$ to $\succsim$ is allowed by the uniqueness of the probability transformation function $\varphi_s(p)$ and the independence of the utility function on s. □

We can note that a decision maker with $\varphi_s(p) \leq p$ systematically underestimates the probabilities of the favorable events and then is called pessimist under risk. Moreover, we obtain the following result.

Corollary 1. *Let $s, s' \in \mathcal{S}$. $\varphi(s,p) \geq \varphi(s',p)$ for any $p \in [0,1]$ if and only $V(s,L) \geq V(s',L)$ for any $L \in \mathcal{L}$.*

Proof. Let $L = (z_1, p_1; z_2, p_2; \ldots; z_n, p_n)$ with $z_1 \leq \ldots \leq z_n$. Note that this axiom implies that for any $z, z' \in \mathcal{Z}$ such that $z \geq z'$, $(s,z) \succsim_s (s,z')$.

(i) $\Rightarrow$: $V(s,L) - V(s',L) = \sum\limits_{i=2}^{n} \left(\varphi_s(\sum\limits_{j=i}^{n} p_j) - \varphi_{s'}(\sum\limits_{j=i}^{n} p_j) \right) \left(v(z^i) - v(z^{i-1}) \right) \geq 0$ if $\varphi(s,p) \geq \varphi(s',p)$ for any $p \in [0,1]$;

(ii) $\Leftarrow$: if there exists p_0 such that $\varphi(s,p_0) < \varphi(s',p_0)$ then for $L_0 = (z_1, 1 - p_0; z_2, p_0), V(s,L_0) < V(s',L_0)$. □

This result implies that if there exists a realized state that induces pessimistic risk perception, then an individual will dislike any decision in this context, with respect to a context where his risk perception is less pessimistic. It is possible in this case to consider that s is preferred to s'.

When the realized state is a past realization of decision relevant event (as loss realizations in insurance decisions), availability bias and gambler fallacy can lead to very different relations between past experience and pessimism. More precisely, an individual prone to the availability bias will become more pessimistic after a loss realization than after a no loss period whereas an individual prone to gambler's fallacy will become less pessimistic after a loss realization than after a no loss period. Then, under availability bias, a period following the occurrence of a loss will be perceived as worst than a period following no loss and under gambler's fallacy, a period following the occurrence of a loss will be perceived as better than a period following no loss.

3 Dynamic Choice

In this section, we consider a dynamic choice problem under risk where risk perception and utility of the outcomes may depend on agents' past experience. Preferences at a point of time are represented as in the previous section of the paper, by a past experience dependent RDU. It is now well known (see for instance Machina (1989)) that preferences representations models that do not verify the independence axiom cannot verify at the same time dynamic consistency, consequentialism and reduction of compound lotteries. To preserve dynamic consistency, as in Epstein and Schneider (2003), Epstein and Wang (1994), Hayashi (2005) and Klibanoff, Marinacci, and Mukerji (2006), a recursive model is adopted here. More precisely, we modify the Kreps and Porteus (1978) model in order to introduce both risk perception and past experience dependence.

3.1 Some Notations and Definitions

We consider a discrete and finite sequence of times $t = 1,...,T$. Z_t is the set of possible payoffs at time t. To simplify, we assume that $Z_t = Z$ which is a compact interval of R for any t from 1 to T. A payoff realized at time t is denoted by z_t. Decision maker past experience at time t is characterized by a sequence of events, relevant for the considered decision and denoted by s_t. More precisely, $s_t = (e_0, e_1, ..., e_t)$ where e_τ is the event that occurred at time τ with $e_\tau \in \mathcal{E}_\tau$, the set of all possible events at time τ and $e_0 \in \mathcal{E}_0$ the set of all possible past experiences. S_t is then the set of possible histories up to time t verifying the recursive relation: $S_0 = \mathcal{E}_0$ and $S_t = S_{t-1} \times \mathcal{E}_t$. We denote by $\mathcal{M}(\mathcal{E}_t)$ the set of distributions on $\mathcal{E}_t$.

At period T, $\mathcal{L}_T$ is the set of distributions on Z_T endowed with the Prohorov metric. X_T, the set of risky perspectives in which agent has to make his choice, is assumed to be the set of closed non empty subsets of $\mathcal{L}_T$, endowed with the Hausdorff metric.

By recurrence, we define, $\mathcal{L}_t$, the set of probability distributions on $\mathcal{C}_t = Z_t \times X_{t+1} \times \mathcal{M}(\mathcal{E}_{t+1})$ with X_{t+1} the set of closed non empty subsets of $\mathcal{L}_{t+1}$.

At each period, the nature chooses a probability distribution on $\mathcal{E}_{t+1}$. The agent cannot influence this distribution. Given this distribution, the agent has to choose a lottery in the set $\mathcal{L}_t$. The assumption of compound lotteries reduction is made between distributions on wealth and events for a fixed period. However, this assumption is relaxed between two consecutive periods.

For each period t, $\succsim_t$ denotes a binary relation on $s_t \times \mathcal{L}_t$ for a given s_t. We denote $\succ_t$ the strict preference and $\sim_t$ the indifference.

Axiom 7 (1bis) *$\succsim_t$ is a complete order on $s_t \times \mathcal{L}_t$.*

We assume that $\succsim_t$ verifies Axioms 2–5 on $s_t \times \mathcal{L}_t$.

Under these previous axioms, we can represent preferences at a point of time in the similar way to Theorem 1:

Lemma 1. *For any s_t in S_t, Axioms 1bis, 2–5 are necessary and sufficient for there to exist, for each t, a bounded continuous function $v_t : s_t \times Z_t \times X_{t+1} \times \mathcal{E}_{t+1} \longrightarrow R$ and a continuous function $\varphi_t : s_t \times [0,1] \longrightarrow [0,1]$ such that for $L_t, L_t' \in \mathcal{L}_t$, $(s_t, L_t) \succsim_t (s_t, L_t')$ if and only if $V_t(s_t, L_t) \geq V_t(s_t, L_t')$ with*

$$V_t(s_t, L_t) = \sum_{i=1}^{n} \left(\varphi_t(s_t, \sum_{j=i}^{n} p_j) - \varphi_t(s_t, \sum_{j=i+1}^{n} p_j) \right) \times v_t(s_t, z_t, x_{t+1}, e_{t+1}) \equiv RDv_t(s_t, L_t).$$

The proof comes immediately from the previous section. Let us notice that utility function v_t depends on past experience s_t whereas in a static problem, v did not depend on state s. This comes from the fact that in a dynamic choice, future perspective depends on past experience. We will precise this point in the next section.

3.2 Temporal Consistency and the Representation Theorem

We adapt the KP temporal consistency axiom to our context in the following manner.

Axiom 8 *(Temporal consistency) We consider the degenerate distributions, Δ_{e_t} in $\mathcal{M}(\mathcal{E}_t)$ and for a given distribution $\delta_{e_{t+1}}$ in $\Delta_{e_{t+1}}$, the degenerate distributions on $Z_t \times X_{t+1} \times \delta_{e_{t+1}}$. Then, for all t, $s_t \in S_t$, $e_{t+1} \in \mathcal{E}_{t+1}$, $z_t \in Z_t$, $x_{t+1}, x'_{t+1} \in X_{t+1}$,*

$(s_t, z_t, x_{t+1}, e_{t+1}) \succsim_t (s_t, z_t, x'_{t+1}, e_{t+1})$ iff $(s_{t+1}, x_{t+1}) \succsim_{t+1} (s_{t+1}, x'_{t+1})$ with $s_{t+1} = (s_t, e_{t+1})$ at period $t+1$.

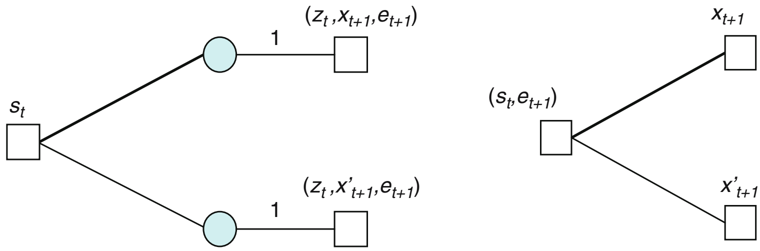


Fig. 1

Let us consider the lotteries in Fig. 1.

The temporal consistency axiom sets that if the degenerate lottery (z_t, x_{t+1}, e_{t+1}) is preferred to the degenerate lottery (z_t, x'_{t+1}, e_{t+1}) for a decision maker with past experience s_t, then when e_{t+1} is realized, the decision maker will choose x_{t+1} between x_{t+1} and x'_{t+1}. In the same way, if at time $t+1$, he chooses x_{t+1}, when e_{t+1} is realized, then he cannot at time t, strictly prefer (z_t, x'_{t+1}, e_{t+1}) to (z_t, x_{t+1}, e_{t+1}).

Lemma 2. *Axioms 1bis, 2–5 and 8 are necessary and sufficient for there to exist functions v_t as in previous lemma and, for fixed $\{v_t\}$, unique functions*
$u_t : \{(s_t, z_t, \gamma) \in S_t \times Z_t \times R : \gamma = RDv_{t+1}(s_{t+1}, L_{t+1})\} \to R$
which are strictly increasing in their third argument and which satisfy
$v_t(s_t, z_t, x_{t+1}, e_{t+1}) = u_t(s_t, z_t, L_{t+1\max} RDv_{t+1}(s_{t+1}, L_{t+1}))$
for all $s_t \in S_t$, $e_{t+1} \in \mathcal{E}_{t+1}$, $z_t \in Z_t$, $x_{t+1} \in X_{t+1}$.

Proof. Axioms 1bis-5 and 8 are hold, Lemma 3 fix $V_t(s_t, L_t)$. Then, $V_{t+1}((s_t, e_{t+1}), L_{t+1}) = V_{t+1}\left((s_t, e_{t+1}), L'_{t+1}\right) \Longrightarrow V_t(s_t, L_{t+1}) = V_t\left(s_t, L'_{t+1}\right)$ for a given e_{t+1} (axiom TC). Consequently, u_t is strictly increasing in its third argument.

(ii) If V_t and u_t are given with u_t is strictly increasing in its third argument, V_t verifies Lemma 3. Then Axioms 1bis-6 hold. u_t is increasing in its third argument then $V_t(s_t, L) \geq V_t(s_t, L') \Longleftrightarrow u_t(s_t, z_t, V_{t+1}(s_{t+1}, L)) \geq u_t(s_t, z_t, V_{t+1}(s_{t+1}, L'))$
$\Longleftrightarrow V_{t+1}((s_t, e_{t+1}), L) \geq V_{t+1}((s_t, e_{t+1}), L')$. Then Axiom 8 holds. □

As we can see, in the dynamic problem, utility functions v_t depend on past experience s_t. Indeed, at time T, utility function does not depend on past experience. But, at time $T-1$, the certainty equivalent of lottery, given by u_{T-1}, depends on past experience $s_{T-1.}$ Consequently, v_{T-1} directly depends on past experience. Recursively, at each period, certainty equivalent depends on past experience and then utility function too.

Theorem 3. *Axioms 1bis, 2–5 and 8 are necessary and sufficient for there to exist a continuous function $v : S_T \times Z_T \to R$ and, for $t = 0, \ldots, T-1$, continuous functions $u_t : S_t \times Z_t \times R \to R$, strictly increasing in their third argument, so that, $v_T(s_T, z_T) = v(z_T)$ and, recursively*
$v_t(s_t, z_t, x_{t+1}, e_{t+1}) = L_{t+1\max} u_t(s_t, z_t, RDv_{t+1}(s_{t+1}, L_{t+1}))$,
then, for all $s_t \in S_t$, $L_t, L'_t \in \mathcal{L}_t$, $(s_t, L_t) \succsim_t (s_t, L'_t)$ iff $RDv_t(s_t, L_t) \geq RDv_t(s_t, L'_t)$

$$with\ RDv_t\left(s_t, L_t\right) = \sum_{i=1}^{n}\left(\varphi_t(s_t, \sum_{j=i}^{n} p_j) - \varphi_t(s_t, \sum_{j=i+1}^{n} p_j)\right) \times v_t(s_t, l_t).$$

Proof. We adapt the proof of the theorem in Kreps and Porteus. □

Note that the preference representation requires v, the functions u_t and the functions φ_t to implicitly define functions v_t. As in Kreps and Porteus, it introduces the concept of timing of resolution of uncertainty. This representation can explain some intertemporal behaviors not explained by the standard Expected Utility model. We propose in the next section an illustration.

4 An Insurance Demand Illustration

In this section, we study the implications of the previous model for multi-period demand decisions on the insurance market. It appears that introducing a relation between realized damages and risk perception gives an explanation for some observed insurance demand patterns against catastrophic risk.

We study the optimal insurance demand strategy of an individual for three periods of time (years). The individual faces a risk of loss of amount L at each period. There exists a perfectly competitive insurance market proposing insurance contracts at a fair premium. Insurance contracts are subscribed for one period. Consequently, the individual has to choose an amount of coverage at each period. We assume that for one period the estimated probability of incurring a loss is p and that losses in successive periods are independent:

P (loss at period t/loss in period $t-1$) = p.

At each period, insurance contracts C_t are proposed. They are characterized by pairs (indemnity I_t, premium Π_t) such that $I_t = \alpha_t L$ with $\alpha_t \in [0,1]$ and $\Pi_t = \alpha_t pL$. At period $t = 0$, the agent receives a certain wealth, z_0, and he is in the state s_0. Past experience is resumed by the sequences of events $\{$damage, no damage$\}$. We denote by e_t the event "damage at period t" and by e_t' the event "no damage at period t". At each period, individual has to choose α_t. The corresponding decision tree is given in Fig. 2.

At a point of time, the probability transformation function is assumed to be the following:

$$\varphi_t\left(p, s_t\right) = p^{\sum_{\tau=0}^{t} e_\tau}$$

with $e_0 = \frac{1}{2}$, $e_t = \frac{1}{2}$ and $e_t' = 0$ for $t = 1,2$.

In this case, the individual is optimistic at period 0 and modifies his risk perception with respect to damages, occurring or not: the occurrence of a damage modifies his risk perception and he becomes less optimistic, if no damages occurs, his risk perception does not change. This kind of risk perception illustrates in some sense the previously mentioned availability bias.

To better isolate the risk perception influence, we assume that preferences at the final period are represented by a linear utility function under certainty.

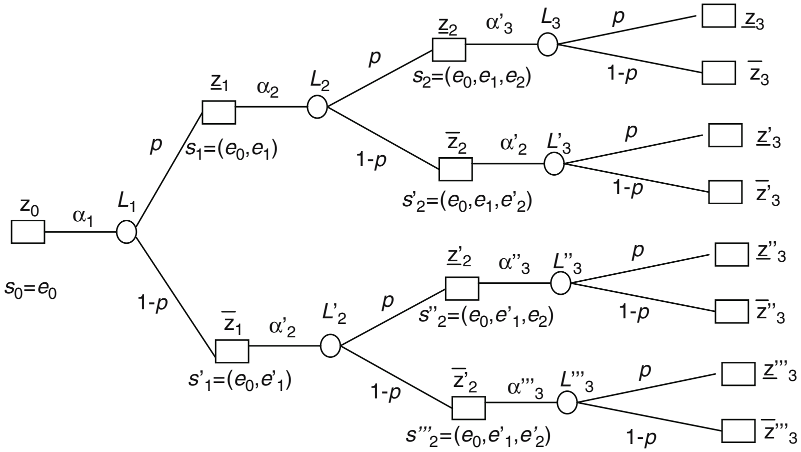


Fig. 2

Moreover, for simplicity, we suppose that the individual is neutral toward the time of resolution of uncertainty (in the sense of Kreps and Porteus), so that

$$u_t\left(s_t, z_t, RDv_{t+1}\left(s_{t+1}, L_{t+1}\right)\right) = RDv_{t+1}\left(s_{t+1}, L_{t+1}\right).$$

The dynamic choice problem solves in several steps. Note that, due to the linear utility assumption, only corner solutions will prevail.

(i) For each terminal node, we have to compute $v(s_3, z_3) = z_3$, with z_3 the wealth at the final period.
(ii) For each final decision node, we have to evaluate the individual utility and maximize it.

For example, at node L_3, for a coverage rate α_3, the utility writes:

$$V_3\left(s_3, L_3\right) = RDv\left(s_3, L_3\right) = v\left(s_3, \underline{z_3}\right) + \left[v\left(s_3, \bar{z}_3\right) - v\left(s_3, \underline{z_3}\right)\right] \times \varphi_3\left(1-p, s_3\right)$$

with $\underline{z_3} = z_2 - \Pi_3 - L + I_3 = z_2 - L + \alpha_3 L(1-p)$, $\bar{z}_3 = z_2 - \Pi_3 = z_2 - \alpha_3 Lp$, and $\varphi_3\left(1-p, s_3\right) = (1-p)^{e_0+e_1+e_2} = (1-p)^{\frac{3}{2}}$.

Then,

$$\begin{aligned} V_3\left(s_3, L_3\right) &= z_2 - L + \alpha_3 L(1-p) + L\left[1-\alpha_3\right] \times (1-p)^{\frac{3}{2}} \\ &= \alpha_3 L(1-p)\left[1-(1-p)^{\frac{1}{2}}\right] + z_2 - L + L(1-p)^{\frac{3}{2}} \end{aligned}$$

As utility is an increasing function of the coverage rate, the optimal coverage is the full coverage and the utility value becomes:

$$V_3\left(s_3, L_3^*\right) = z_2 - pL \tag{1}$$

In the same way, we obtain that $V_3\left(s_3', L_3'^{*}\right) = z_2 - pL$ for any $\alpha_3' \in [0,1]$, $V_3\left(s_3'', L_3''^{*}\right) = z_2 - pL$ for any $\alpha_3'' \in [0,1]$ and $V_3\left(s_3''', L_3'''^{*}\right) = z_2 - L\left(1-(1-p)^{1/2}\right)$ for $\alpha_3'''^{*} = 0$.

(iii) At period 2, we have four nodes which values are:
$u_2\left(s_2, \underline{z}_2, RDv_3\right) = \underline{z}_2 - pL$ with $\underline{z}_2 = z_1 - \Pi_2 - L + I_2$,
$u_2\left(s_2, \overline{z}_2, RDv_3\right) = \overline{z}_2 - pL$ with $\overline{z}_2 = z_1 - \Pi_2$,
$u_2\left(s_2, \underline{z}_2', RDv_3\right) = \underline{z}_2' - pL$ with $\underline{z}_2' = z_1 - \Pi_2' - L + I_2'$ and
$u_2\left(s_2, \overline{z}_2', RDv_3\right) = \overline{z}_2' - L\left(1-(1-p)^{1/2}\right)$ with $\overline{z}_2' = z_1 - \Pi_2'$.

(iv) We repeat step (ii). Then, at node L_2, for a coverage rate α_2, the utility writes:

$$V_2\left(s_2, L_2\right) = RDv_2\left(s_2, L_2\right) = v_2\left(s_2, \underline{z_2}\right) + \left[v_2\left(s_2, \overline{z}_2\right) - v_2\left(s_2, \underline{z_2}\right)\right] \times \varphi_2\left(1-p, s_2\right)$$

with $v_2\left(s_2, \underline{z_2}\right) = u_2\left(s_2, \underline{z}_2, RDv_3\right) = \underline{z}_2 - pL = z_1 - L(1+p) + \alpha_2 L(1-p)$, $v_2\left(s_2, \overline{z}_2\right) = u_2\left(s_2, \overline{z}_2, RDv_3\right) = \overline{z}_2 - pL = z_1 - pL - \alpha_2 pL$ and $\varphi_2\left(1-p, s_2\right) = (1-p)^{e_0+e_1} = 1-p$.

Then,

$$V_2\left(s_2, L_2\right) = z_1 - 2Lp$$

for any $\alpha_2 \in [0,1]$.

At node L_2', we have to pay attention to the value of $v_2\left(s_2', \underline{z}_2'\right) = u_2\left(s_2', \underline{z}_2', RDv_3\right)$ and $v_2\left(s_2', \overline{z}_2'\right) = u_2\left(s_2', \overline{z}_2', RDv_3\right)$ since, in the RDU framework, we have to rank utility.

In our example, $v_2\left(s_2', \underline{z}_2'\right) < v_2\left(s_2', \overline{z}_2'\right)$. Thus, we obtain that

$$V_2\left(s_2', L_2'\right) = RDv_2\left(s_2', L_2'\right) = v_2\left(s_2', \underline{z_2'}\right) + \left[v_2\left(s_2', \overline{z}_2'\right) - v_2\left(s_2', \underline{z_2'}\right)\right] \times \varphi_2\left(1-p, s_2'\right)$$

with $v_2\left(s_2', \underline{z_2'}\right) = \underline{z}'_2 - pL = z_1 - L(1+p) + \alpha_2' L(1-p)$, $v_2\left(s_2', \overline{z}_2'\right) = \overline{z}_2' - L\left(1-(1-p)^{1/2}\right) = z_1 - L\left(1-(1-p)^{1/2}\right) - \alpha_2' pL$ and $\varphi_2\left(1-p, s_2'\right) = (1-p)^{e_0+e_1'} = (1-p)^{1/2}$.

Then,

$$V_2\left(s_2', L_2'\right) = z_1 + Lp\left[(1-p)^{1/2} - 2\right] + \alpha_2' L (1-p)^{1/2}\left[(1-p)^{1/2} - 1\right]$$

As utility is a decreasing function of the coverage rate, the optimal coverage is null and the value of utility becomes

$$V_2\left(s_2', L_2'^{*}\right) = z_1 + Lp\left[(1-p)^{1/2} - 2\right] \tag{2}$$

(v) At period 1, we have two nodes which values are:
$u_1\left(s_1, \underline{z}_1, RDv_2\right) = \underline{z}_1 - 2pL$ with $\underline{z}_1 = z_0 - \Pi_1 - L + I_1$ and
$u_1\left(s_1, \overline{z}_1, RDv_2\right) = \overline{z}_1 + Lp\left[(1-p)^{1/2} - 2\right]$ with $\overline{z}_1 = z_0 - \Pi_1$.

Then, at node L_1, for a coverage rate α_1, the utility writes:

$$V_1(s_1,L_1) = RDv_1(s_1,L_1) = v_1\left(s_1,\underline{z_1}\right) + \left[v_1(s_1,\bar{z}_1) - v_1\left(s_1,\underline{z_1}\right)\right] \times \varphi_1(1-p,s_1)$$

with $v_1\left(s_1,\underline{z_1}\right) = \underline{z}_1 - 2pL = z_0 - \Pi_1 - L + I_1 - 2pL = \alpha_1 L(1-p) + z_0 - L(1+2p)$, $v_1(s_1,\bar{z}_1) = \bar{z}_1 + Lp\left[(1-p)^{1/2} - 2\right] = z_0 - \alpha_1 pL + Lp\left[(1-p)^{1/2} - 2\right]$ and $\varphi_1(1-p,s_1) = (1-p)^{e_0} = (1-p)^{1/2}$.

Then,

$$V_1(s_1,L_1) = \alpha_1 L(1-p)^{1/2}\left[(1-p)^{1/2} - 1\right] + z_0 - L(1+2p) + L\left[p(1-p)^{1/2} + 1\right](1-p)^{1/2}$$

As utility is a decreasing function of the coverage rate, the optimal coverage is zero and the value of utility becomes

$$V_1(s_1,L_1^*) = z_0 - L(1+2p) + L\left[p(1-p)^{1/2} + 1\right](1-p)^{1/2} \tag{3}$$

To summarize, the results are the following:

- $\alpha_1 = 0$;
- $\alpha_2 = 0$ if no loss at period 1;
  $\alpha_2 \in [0,1]$ if loss at period 1;
- $\alpha_3 = 0$ if no loss at periods 1 and 2;
  $\alpha_3 = 1$ if loss at periods 1 and 2;
  $\alpha_3 \in [0,1]$ else.

In this illustration, the individual chooses not to buy insurance in the first period. In the second period, he chooses not to be covered only if he had not damage. Finally, in the third period, the two extreme insurance coverage decisions are possible: if the individual had never incurred a loss, he chooses not to buy insurance; if he had two consecutive losses, he buys full coverage and in the intermediate cases, he is indifferent between all the insurance levels.

This example underlines the fact that what is important for the decision maker is not only the event occurring in the period directly preceding the moment of the decision, but the all sequence of events, that is all the past experience.

Let us now compare the predictions of our model with those of some standard models:

- The particular case when $\varphi_t(p,s_t) = p$: this corresponds to the standard version of a recursive expected utility model.
  The results are then the following: $\alpha_i \in [0,1]$ for any $i = 1,2,3$. The individual is indifferent between different amounts of coverage and this, at any period and for any experienced damage.

- The resolute choice model proposed by McClennen (1990). In this non consequentialist model, all the strategies are evaluated at the root node and compared according to the root node preferences. We consider two cases: the case when one-shot preferences are EU and the case when one-shot preferences are RDU. The results are then the following:
  (1) For EU preferences, $\alpha_i \in [0,1]$ for any $i = 1,2,3$. The individual is indifferent between different amounts of coverage and this, at any period and for any experienced damage.
  (2) For RDU preferences, $\alpha_i = 0$ or 1 for any i.
  – For $\varphi(p) < p$, complete coverage at any period and for any experienced loss is preferred to a strategy consisting in buying insurance only after experiencing a loss
  – For $\varphi(p) > p$, no coverage at any period and for any experienced loss is preferred to a strategy consisting in buying insurance only after experiencing a loss

5 Concluding Remarks

The insurance demand example shows that our model allows to explain the modifications in the insurance demand behavior over time observed for catastrophic risk and given in the introduction. It well appears that past experience have a cumulative effect on decisions: an individual can maintain constant its insurance demand after one occurrence of the loss and modify it only after two, or more consecutive loss events. In this example, we assumed that at any period observing a loss renders the individual more pessimistic. This explains a behavior in accordance with the availability bias. The gambler's fallacy attitude could be explained if the individual becomes more and more optimistic after experiencing losses.

The comparison with other models shows that neither the recursive model alone, nor the RDU model alone can explain all the observed pattern of behavior.

The insurance example corresponds to the particular case when past experience (context) is composed by the decision-relevant events. Considering different events, that do not directly influence the outcomes (as weather condition in investment decisions) will make even easier to underline the new insights of the present model because of the complete absence of these events in the preferences representations of the standard models.

Acknowledgements We thank A. Chateauneuf, R. Giraud, J.Y. Jaffray, E. Karni, P. Klibanoff, J.C. Vergnaud and the participants of FUR XII and of the research program "Risk Attitude" for valuable comments and useful suggestions.

References

Browne, M., & Hoyt, R. (2000). The demand for flood insurance: Empirical evidence. *Journal of Risk and Uncertainty, 20*, 291–306

Caplin, A., & Leahy, J. (2001). Psychological expected utility theory and anticipatory feeling. *The Quarterly Journal of Economics, 116*, 55–79

Chateauneuf, A. (1999). Comonotonicity axioms and rank-dependent expected utility theory for arbitrary consequences. *Journal of Mathematical Economics, 32*, 21–45

Epstein, L., & Schneider, M. (2003). Recursive multiple priors. *Journal of Economic Theory, 113*, 1–31

Epstein, L., & Wang, T. (1994). Intertemporal asset pricing under Knightian uncertainty. *Econometrica, 62*, 283–322

Ganderton, P., Brookshire, D., McKee, M., Stewart, S., & Thurston, H. (2000). Buying insurance for disaster-type risk: Experimental evidence. *Journal of Risk and Uncertainty, 20*(3), 271–289

Hayashi, T. (2005). Intertemporal substitution, risk aversion and ambiguity aversion. *Economic Theory, 25*(4), 933–956

Hirshleifer D., & Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *Journal of Finance, 58*(3), 1009–1032

Karni, E. (1985). *Decision making under uncertainty: The case of state-dependant preferences*. Cambridge: Harvard University Press

Klibanoff, P., Marinacci, M., & Mukerji, S. (2006). Recursive smooth ambiguity preferences. *Carlo Alberto Notebooks*, No. 17

Kreps, D., & Porteus, E. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica, 46*(1), 185–200

Kunreuther, H. (1996). Mitigation disaster losses through insurance. *Journal of Risk and Uncertainty, 12*, 171–187

Machina, M. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Litterature, 27*, 1622–1668

McClelland, G., Schulze, W., & Coursey D. (1993). Insurance for low-probability hazards: A bimodal response to unlikely events. *Journal of Risk and Uncertainty, 7*, 95–116

McClennen, E. (1990). *Rationality and dynamic choice*. Cambridge: Cambridge University Press

Papon, T. (2004). The effect of precommitment and past-experience on insurance choices: An experimental study. *Cahiers de la MSE*, No. 40

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization, 3*(4), 323–343

Slovic, P. (2000). *The perception of risk*. London and Sterling: Earthscan Publications Ltd

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207–232

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59*, 251–278

Yaari, M. (1987). The dual theory of choice under risk. *Econometrica, 55*, 95–105

Representation of Conditional Preferences Under Uncertainty

G.B. Asheim

Keywords: Decision making under uncertainty · Representation results · Game theory

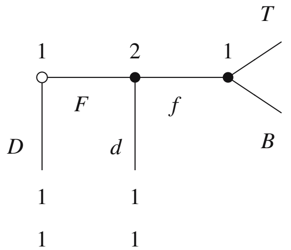
1 Introduction

The purpose of this paper is to present axioms for systems of conditional preferences, and to provide representation results for various sets of such axioms. The preferences are over acts, being functions from uncertain states to outcomes, while the representations are in terms of beliefs over the states and utility of the outcomes. It builds on Asheim (2006, Chap. 3) and Asheim and Perea (2005).

The central question posed by the present paper is the following: *Does a preference between two acts change when new information, ruling out states at which the two acts lead to the same outcomes, becomes available?* At first thought one may conclude that such news are uninformative about the relative merits of the two acts and hence should not influence the preference of the decision maker. However, by considering Kreps and Wilson's (1982) *sequential equilibrium* in game Γ_1 of Fig. 1, we are lead towards a different conclusion. Assume that both players' payoffs are less than one if F and f are chosen. Still, the play of F by player 1 is part of a sequential equilibrium where 1 believes that 2 makes his rational choice of playing d. This means that player 1 unconditionally considers all three of his strategies as equally good, yielding him a payoff of one. However, if by surprise player 2 responds to F by playing f, then player 1's preference between FT and FB will be refined (provided that FT and FB yield different payoffs for 1 when 2 chooses f). Treating player 1 as the decision maker, player 2's strategies d and f as the uncertain

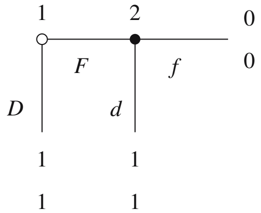
G.B. Asheim
Department of Economics, University of Oslo, P.O. Box 1095 Blindern, NO-0317 Oslo, Norway
e-mail: g.b.asheim@econ.uio.no

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.
© Springer-Verlag Berlin Heidelberg 2008



	d	f
FT	1,1	−,−
FB	1,1	−,−
D	1,1	1,1

Fig. 1 Γ_1 and its strategic form



	d	f
F	1,1	0,0
D	1,1	1,1

Fig. 2 Γ_2 and its strategic form

states he faces, and FT and FB as two acts, we see that 1's preference between these two acts change when state d is ruled out, even though FT and FB lead to the same outcome in this state.

In Asheim and Perea (2005) we epistemically characterize sequential equilibrium and van Damme's (1984) *quasi-perfect equilibrium*, and define the corresponding non-equilibrium concepts – *sequential rationalizability* and *quasi-perfect rationalizability*. These characterizations require that conditional preferences be specified.

How can conditional preferences be specified? There are various ways to do so. One possibility is to represent a system of conditional preferences by means of a *conditional probability system* (CPS) where each conditional belief is a subjective probability distribution.[1] Another possibility is to apply a single sequence of subjective probability distributions – a so-called *lexicographic probability system* (LPS) as defined by Blume, Brandenburger, and Dekel (1991) – and derive the conditional beliefs as the conditionals of such an LPS. Since each conditional LPS is found by constructing a new sequence, which includes the well-defined conditional probability distributions of the original sequence, each conditional belief is itself an LPS.

However, the analysis of Asheim and Perea (2005) shows that neither a CPS nor a single LPS may adequately represent a system of conditional preferences.

Quasi-perfectness cannot always be modeled by a CPS since the modeling of preference for cautious behavior may require lexicographic probabilities. To see this, consider game Γ_2 of Fig. 2. In this game, if player 1 believes that player 2 chooses rationally, then player 1 must assign probability one to player 2 choosing d. Hence, if each (conditional) belief is associated with a subjective probability distribution – as is the case with the concept of a CPS – and player 1 believes that his opponent chooses rationally, then player 1 is indifferent between his two strategies. This is inconsistent with quasi-perfectness, which requires players to have preference for cautious behavior, meaning that player 1 in Γ_2 prefers D to F.

Moreover, sequentiality cannot always be modeled by means of conditionals of a single LPS since preference for cautious behavior is induced. To see this, consider a modified version of Γ_2 where an additional subgame is substituted for the $(0, 0)$-payoff, with all payoffs in that subgame being smaller than one. If player 1's conditional beliefs over strategies for player 2 is derived from a single LPS, then a well-defined belief conditional on reaching the added subgame entails that player 1 deems possible the event that player 2 chooses f, and hence, player 1 prefers D to F. This is inconsistent with sequentiality, under which F is a rational choice.

Therefore, this paper reports on a new way of representing a system of conditional preferences, by means of what Andrés Perea and I call a *system of conditional lexicographic probabilities* (SCLP) (cf. Asheim & Perea, 2005). In contrast to a CPS, an SCLP may induce conditional beliefs that are represented by LPSs rather than subjective probability distributions. In contrast to the system of conditionals derived from a single LPS, an SCLP need not include all levels in the sequence of the original LPS when determining conditional beliefs. Thus, an SCLP ensures well-defined conditional beliefs representing nontrivial conditional preferences, while allowing for flexibility w.r.t. whether to assume preference for cautious behavior. This is accomplished by combining an LPS with a length function, specifying for each conditioning event the number of levels of the original LPS used to represent the conditional preferences.

The analysis is based on the Anscombe–Aumann framework (Anscombe & Aumann, 1963), where preferences are defined over functions from states to *objective randomizations* over outcomes. Such functions will be referred to as *Anscombe–Aumann acts* (in contrast to acts in the Savage (1954), sense; i.e., functions from states to deterministic outcomes). A strategy in a game is a function that for each opponent strategy choice, determines an outcome. A pure strategy determines for each opponent strategy a deterministic outcome, while a mixed strategy determines for each opponent strategy an objective randomization over the set of outcomes. Hence, a pure strategy is an example of an act in the sense of Savage (1954), while a mixed strategy is an example of an act in the generalized sense of Anscombe and Aumann (1963).

Allowing for objective randomizations and using Anscombe–Aumann acts are convenient in game-theoretic applications for the following reason: The Anscombe–Aumann framework allows a player's payoff function to be a von Neumann-Morgenstern (vNM) utility function determined from his preferences over randomized outcomes, independently of the likelihood that he assigns to the different strategies of his opponent. This is consistent with the way games are normally presented, where payoff functions for each player are provided independently of the analysis of the strategic interaction.

In addition to allowing for flexibility concerning how to specify conditional preferences, the Anscombe–Aumann framework will be generalized in two other ways in this paper.

First, I will follow Blume et al. (1991) by imposing the *conditional Archimedean property* (also called *conditional continuity*) instead of *Archimedean property* (also called *continuity*). This is important for modeling caution, which requires a player to take into account the possibility that the opponent makes an irrational choice, while assigning probability one to the event that the opponent makes a rational choice. That is, even though any irrational choice is infinitely less likely than some rational choice, it is not ruled out. Such discontinuous preferences are also useful when modeling players' preferences in extensive games.

Second, I will relax the axiom of *completeness* to *conditional completeness*. While complete preferences will normally be represented by means of subjective probabilities (cf. Propositions 1, 2, 3, and 5 of this paper), incomplete preferences are insufficient to determine the relative likelihood of the uncertain states.[2] Subjective probabilities are not part of the most common deductive procedures in game theory – like iterated elimination of strongly dominated strategies, the Dekel–Fudenberg procedure, and the backward induction procedure. One can argue that, since they make no use of subjective probabilities, one should seek to provide epistemic conditions for such procedures without reference to subjective probabilities (see, e.g., Asheim, 2006, Chaps. 6 and 7). Indeed, subjective probabilities play no role in the epistemic analysis of backward induction by Aumann (1995). Moreover, forward induction can fruitfully be modeled by means of incomplete preferences (cf. Asheim & Dufwenberg, 2003).

Proofs of the results reported below are contained in an appendix.

2 Axioms

Consider a decision maker under uncertainty, and let F be a finite set of *states*. The decision maker is uncertain about what state in F will be realized. Let Z be a finite set of *outcomes*. For each $\varphi \in 2^F\backslash\{\emptyset\}$, the decision maker is endowed with a binary relation (*preferences*) over all functions that to each element of φ assign an objective randomization on Z. Any such function is called an *act* on φ, and is the subject of analysis in the decision-theoretic framework introduced by Anscombe and Aumann (1963). Write $\mathbf{p}_\varphi$ and $\mathbf{q}_\varphi$ for acts on $\varphi \in 2^F\backslash\{\emptyset\}$. (For acts on F, write simply $\mathbf{p}$ and $\mathbf{q}$.) A binary relation on the set of acts on φ is denoted by $\succsim_\varphi$, where $\mathbf{p}_\varphi \succsim_\varphi \mathbf{q}_\varphi$ means that $\mathbf{p}_\varphi$ is *preferred* or *indifferent* to $\mathbf{q}_\varphi$. As usual, let $\succ_\varphi$ (*preferred to*) and $\sim_\varphi$ (*indifferent to*) denote the asymmetric and symmetric parts of $\succsim_\varphi$.

Consider the following five axioms, where the numbering of axioms follows Blume et al. (1991).

Axiom 1 (Order) $\succsim_\varphi$ *is complete and transitive.*

Axiom 2 (Objective Independence) $\mathbf{p}'_\varphi \succ_\varphi$ *(resp.* $\sim_\varphi$*)* $\mathbf{p}''_\varphi$ *iff* $\gamma\mathbf{p}'_\varphi + (1-\gamma)\mathbf{q}_\varphi \succ_\varphi$ *(resp.* $\sim_\varphi$*)* $\gamma\mathbf{p}''_\varphi + (1-\gamma)\mathbf{q}_\varphi$*, whenever* $0 < \gamma < 1$ *and* $\mathbf{q}_\varphi$ *is arbitrary.*

Axiom 3 (Nontriviality) *There exist* $\mathbf{p}_\varphi$ *and* $\mathbf{q}_\varphi$ *such that* $\mathbf{p}_\varphi \succ_\varphi \mathbf{q}_\varphi$.

Axiom 4 (Archimedean Property) *If* $\mathbf{p}'_\varphi \succ_\varphi \mathbf{q}_\varphi \succ_\varphi \mathbf{p}''_\varphi$, *then* $\exists 0 < \gamma < \delta < 1$ *such that* $\delta \mathbf{p}'_\varphi + (1-\delta)\mathbf{p}''_\varphi \succ_\varphi \mathbf{q}_\varphi \succ_\varphi \gamma \mathbf{p}'_\varphi + (1-\gamma)\mathbf{p}''_\varphi$.

Say that $e \in F$ is *Savage-null* if $\mathbf{p}_{\{e\}} \sim_{\{e\}} \mathbf{q}_{\{e\}}$ for all acts $\mathbf{p}_{\{e\}}$ and $\mathbf{q}_{\{e\}}$ on $\{e\}$. Denote by κ the non-empty set of states that are *not* Savage-null; i.e., the set of states that the decision maker deems subjectively possible. Write $\Phi := \{\varphi \in 2^F \backslash \{\emptyset\} \mid \kappa \cap \varphi \neq \emptyset\}$. Refer to the collection $\{\succsim_\varphi \mid \varphi \in \Phi\}$ as a *system of conditional preferences* on the collection of sets of acts from subsets of F to outcomes.

Whenever $\emptyset \neq \varepsilon \subseteq \varphi$, denote by $\mathbf{p}_\varepsilon$ the restriction of $\mathbf{p}_\varphi$ to ε.

Axiom 5 (Non-null State Independence) $\mathbf{p}_{\{e\}} \succ_{\{e\}} \mathbf{q}_{\{e\}}$ *iff* $\mathbf{p}_{\{f\}} \succ_{\{f\}} \mathbf{q}_{\{f\}}$, *whenever* $e, f \in \kappa$, *and* $\mathbf{p}_{\{e,f\}}$ *and* $\mathbf{q}_{\{e,f\}}$ *satisfy* $\mathbf{p}_{\{e,f\}}(e) = \mathbf{p}_{\{e,f\}}(f)$ *and* $\mathbf{q}_{\{e,f\}}(e) = \mathbf{q}_{\{e,f\}}(f)$.

Define the *conditional* binary relation of $\succsim_\varphi$ on ε, $\succsim_{\varphi|\varepsilon}$, by $\mathbf{p}'_\varphi \succsim_{\varphi|\varepsilon} \mathbf{p}''_\varphi$ if, for some $\mathbf{q}_\varphi$, $(\mathbf{p}'_\varepsilon, \mathbf{q}_{\varphi\backslash\varepsilon}) \succsim_\varphi (\mathbf{p}''_\varepsilon, \mathbf{q}_{\varphi\backslash\varepsilon})$, where $\emptyset \neq \varepsilon \subseteq \varphi$. By Axioms 1 and 2, this definition does not depend on $\mathbf{q}_\varphi$. Note that $\succsim_{\varphi|\varepsilon}$ corresponds to the decision maker's preferences when having received information that the true state is in φ and only considering the event that the true state is in $\varepsilon \subseteq \varphi$, while $\succsim_\varepsilon$ denotes the decision maker's preferences when having received information that the true state is in ε. The following axiom states that preferences over acts when having received information that the true state is in ε equals the preferences when having received information that the true state is in φ ($\supseteq \varepsilon$) and only considering the event that the true state is in ε.

Axiom 6 (Conditionality) $\mathbf{p}_\varepsilon \succ_\varepsilon$ *(resp.* $\sim_\varepsilon$*)* $\mathbf{q}_\varepsilon$ *iff* $\mathbf{p}_\varphi \succ_{\varphi|\varepsilon}$ *(resp.* $\sim_{\varphi|\varepsilon}$*)* $\mathbf{q}_\varphi$, *whenever* $\emptyset \neq \varepsilon \subseteq \varphi$.

It is an immediate observation that Axioms 5 and 6 imply *non-null state independence* as stated in Axiom 5 of Blume et al. (1991).

Lemma 1 *Assume that the system of conditional preferences* $\{\succsim_\varphi \mid \varphi \in \Phi\}$ *satisfies Axioms 5 and 6. Then,* $\forall \varphi \in \Phi$, $\mathbf{p}_\varphi \succ_{\varphi|\{e\}} \mathbf{q}_\varphi$ *iff* $\mathbf{p}_\varphi \succ_{\varphi|\{f\}} \mathbf{q}_\varphi$ *whenever* $e, f \in \kappa \cap \varphi$, *and* $\mathbf{p}_\varphi$ *and* $\mathbf{q}_\varphi$ *satisfy* $\mathbf{p}_\varphi(e) = \mathbf{p}_\varphi(f)$ *and* $\mathbf{q}_\varphi(e) = \mathbf{q}_\varphi(f)$.

Turn now the relaxation of Axioms 1, 4, and 6, as motivated in the previous section.

Axiom 1′ (Conditional Order) $\succsim_\varphi$ *is reflexive and transitive and,* $\forall e \in \varphi$, $\succsim_{\varphi|\{e\}}$ *is complete.*

Axiom 4′ (Conditional Archimedean Property) $\forall e \in \varphi$, *if* $\mathbf{p}'_\varphi \succ_{\varphi|\{e\}} \mathbf{q}_\varphi \succ_{\varphi|\{e\}} \mathbf{p}''_\varphi$, *then* $\exists 0 < \gamma < \delta < 1$ *such that* $\delta \mathbf{p}'_\varphi + (1-\delta)\mathbf{p}''_\varphi \succ_{\varphi|\{e\}} \mathbf{q}_\varphi \succ_{\varphi|\{e\}} \gamma \mathbf{p}'_\varphi + (1-\gamma)\mathbf{p}''_\varphi$.

Axiom 6′ (Dynamic Consistency) $\mathbf{p}_\varepsilon \succ_\varepsilon \mathbf{q}_\varepsilon$ *whenever* $\mathbf{p}_\varphi \succ_{\varphi|\varepsilon} \mathbf{q}_\varphi$ *and* $\emptyset \neq \varepsilon \subseteq \varphi$.

Axiom 1′ constitutes a weakening of Axioms 1, since completeness implies reflexivity. This weakening is substantive since, in the terminology of Anscombe and

Aumann (1963), it means that the decision maker has complete preferences over 'roulette lotteries' where objective probabilities are exogenously given, but not necessarily complete preferences over 'horse lotteries' where subjective probabilities, if determined, are endogenously derived from the preferences of the decision maker.

Axiom 4$'$ is the weakening of the Archimedean property (Axiom 4) introduced by Blume et al. (1991). Blume et al. (1991) also considers an axiom in between Axioms 4 and 4$'$ (with the numbering also for this axiom following their scheme).

Axiom 4$''$ (Partitional Archimedean Property) *There is a partition* $\{\pi'_1,\ldots,\pi'_{L|\varphi}\}$ *of* $\kappa \cap \varphi$ *such that*

- $\forall \ell \in \{1,\ldots,L|\varphi\}$, *if* $\mathbf{p}'_\varphi \succ_{\varphi|\pi'_\ell} \mathbf{q}_\varphi \succ_{\varphi|\pi'_\ell} \mathbf{p}''_\varphi$, *then* $\exists 0 < \gamma < \delta < 1$ *such that* $\delta \mathbf{p}'_\varphi + (1-\delta)\mathbf{p}''_\varphi \succ_{\varphi|\pi'_\ell} \mathbf{q}_\varphi \succ_{\varphi|\pi'_\ell} \gamma \mathbf{p}'_\varphi + (1-\gamma)\mathbf{p}''_\varphi$, *and*
- $\forall \ell \in \{1,\ldots,L|\varphi - 1\}$, $\mathbf{p}_\varphi \succ_{\varphi|\pi'_\ell} \mathbf{q}_\varphi$ *implies* $\mathbf{p}_\varphi \succ_{\varphi|\pi'_\ell \cup \pi'_{\ell+1}} \mathbf{q}_\varphi$.

Axiom 6$'$ entails that preferences over acts when having received information that the true state is in ε refines the preferences when having received information that the true state is in φ ($\supseteq \varepsilon$) and only considering the event that the true state is in ε.

Say that $e \in \kappa$ is deemed *infinitely more likely* than $f \in F$ (and write $e \gg f$) if $\mathbf{p}_{\{e,f\}} \succ_{\{e,f\}} \mathbf{q}_{\{e,f\}}$ whenever $\mathbf{p}_{\{e\}} \succ_{\{e\}} \mathbf{q}_{\{e\}}$. Consider the following two auxiliary axioms.

Axiom 11 (Partitional Priority) *If* $e' \gg e''$, *then* $\forall f \in F$, $e' \gg f$ *or* $f \gg e''$.

Axiom 16 (Compatibility) *There exists a binary relation* $\succsim^*_F$ *satisfying Axioms 1, 2, and 4$'$ such that* $\mathbf{p} \succ^*_{F|\varphi} \mathbf{q}$ *whenever* $\mathbf{p}_\varphi \succ_\varphi \mathbf{q}_\varphi$ *and* $\emptyset \neq \varphi \subseteq F$.

While it is straightforward that Axiom 1 implies Axiom 1$'$, Axiom 4 implies Axiom 4$'$, and Axiom 6 implies Axiom 6$'$, it is less obvious that

- Axiom 1 together with Axioms 2, 4$'$, 5, and 6 imply Axiom 11
- Axiom 6 together with Axioms 1, 2, 4$'$, and 5, imply Axiom 16

This is demonstrated by the following lemma.

Lemma 2 *Assume that (a)* $\succsim_\varphi$ *satisfies Axioms 1, 2, and 4$'$ if* $\varphi \in 2^F \backslash \{\emptyset\}$, *and (b) the system of conditional preferences* $\{\succsim_\varphi \mid \varphi \in \Phi\}$ *satisfies Axioms 5 and 6. Then* $\{\succsim_\varphi \mid \varphi \in \Phi\}$ *satisfies Axioms 11 and 16.*

Hence, even though Axiom 16 may not express a compelling intuition and is designed to ensure the existence of an underlying LPS in the representation results of the next section (cf. Proposition 5 and Corollary 1), it is implied by the set of axioms that Blume et al. (1991) employ.

Table 1 illustrates the relationships between the sets of axioms that we will consider. The arrows indicate that one set of axioms implies another. The figure indicates what kind of representations the different sets of axioms correspond to, as reported in the next section.

Table 1 Relationships between different sets of axioms and their representations

Complete and continuous	1 2 3 4 5 6 *Prob. distr.*	→	1 2 3 4 5 6′ 16 *CPS*
	↓		
Complete and partitionally continuous	1 2 3 4″ 5 6 *LCPS*		↓
	↓		
Complete and discontinuous	1 2 3 4′ 5 6 *LPS*	→	1 2 3 4′ 5 6′ 16 *SCLP*
	↓		
Incomplete and discontinuous	1′ 11 2 3 4′ 5 6		
	Conditionality		Dynamic consistency

3 Representation Results

In view of Lemma 1 and using the characterization result of Anscombe and Aumann (1963), we obtain the following result under Axioms 1, 2, 3, 4, 5, and 6; cf. Blume et al. (1991, Theorem 2.1).

For the statement of this and later results, denote by $\upsilon : Z \to \mathbb{R}$ a vNM utility function, and abuse notation slightly by writing $\upsilon(p) = \sum_{z \in Z} p(z)\upsilon(z)$ whenever $p \in \Delta(Z)$ is an objective randomization. In this and later results, υ is unique up to positive affine transformations.

Proposition 1 (Anscombe & Aumann, 1963) *The following two statements are equivalent.*

1. *(a) $\succsim_\varphi$ satisfies Axioms 1, 2, and 4 if $\varphi \in 2^F \backslash \{\emptyset\}$, and Axiom 3 if and only if $\varphi \in \Phi$, and (b) the system of conditional preferences $\{\succsim_\varphi | \ \varphi \in \Phi\}$ satisfies Axioms 5 and 6.*
2. *There exist a vNM utility function $\upsilon : \Delta(Z) \to \mathbb{R}$ and a unique subjective probability distribution μ on F with support κ such that, $\forall \varphi \in \Phi$,*

$$\mathbf{p}_\varphi \succsim_\varphi \mathbf{q}_\varphi \ \textit{ iff } \ \sum_{e \in \varphi} \mu|_\varphi(e)\upsilon(\mathbf{p}_\varphi(e)) \geq \sum_{e \in \varphi} \mu|_\varphi(e)\upsilon(\mathbf{q}_\varphi(e)),$$

where $\mu|_\varphi$ is the conditional of μ on φ.

In view of Lemma 1, and using Blume et al. (1991, Theorem 3.1), we obtain the following result under Axioms 1, 2, 3, 4′, 5, and 6.

For the statement of this and later results, we need to introduce formally the concept of a lexicographic probability system. A *lexicographic probability system* (LPS) consists of L levels of subjective probability distributions: If $L \geq 1$ and, $\forall \ell \in \{1, \ldots, L\}$, $\mu_\ell \in \Delta(F)$, then $\lambda = (\mu_1, \ldots, \mu_L)$ is an LPS on F. Denote by $\mathbf{L}\Delta(F)$ the set of LPSs on F. Write $\operatorname{supp}\lambda := \cup_{\ell=1}^{L} \operatorname{supp}\mu_\ell$ for the support of λ. If $\operatorname{supp}\lambda \cap \varphi \neq \emptyset$, denote by $\lambda|_\varphi = (\mu'_1, \ldots \mu'_{L|\varphi})$ the conditional of λ on φ.[3]

Furthermore, for two utility vectors v and w, denote by $v \geq_L w$ that, whenever $w_\ell > v_\ell$, there exists $k < \ell$ such that $v_k > w_k$, and let $>_L$ and $=_L$ denote the asymmetric and symmetric parts, respectively.

Proposition 2 (Blume et al., 1991) *The following two statements are equivalent.*

1. *(a) $\succsim_\varphi$ satisfies Axioms 1, 2, and 4′ if $\varphi \in 2^F \backslash \{\emptyset\}$, and Axiom 3 if and only if $\varphi \in \Phi$, and (b) the system of conditional preferences $\{\succsim_\varphi | \ \varphi \in \Phi\}$ satisfies Axioms 5 and 6.*
2. *There exist a vNM utility function $\upsilon : \Delta(Z) \to \mathbb{R}$ and an LPS λ on F with support κ such that, $\forall \varphi \in \Phi$,*

$$\mathbf{p}_\varphi \succsim_\varphi \mathbf{q}_\varphi \textit{ iff } \Big(\sum\nolimits_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{p}_\varphi(e))\Big)_{\ell=1}^{L|\varphi} \geq_L \Big(\sum\nolimits_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{q}_\varphi(e))\Big)_{\ell=1}^{L|\varphi},$$

where $\lambda|_\varphi = (\mu'_1, \dots \mu'_{L|\varphi})$ is the conditional of λ on φ.

In view of Lemma 1 and using Blume et al. (1991, Theorem 5.3), we obtain the following result under Axioms 1, 2, 3, 4″, 5, and 6.

For the statement of this results, we need to introduce the concept that is called a lexicographic conditional probability system in the terminology of Blume et al. (1991, Definition 5.2). A *lexicographic conditional probability system* (LCPS) consists of L levels of *non-overlapping* subjective probability distributions: If $\lambda = (\mu_1, \dots, \mu_L)$ is an LPS on F and the supports of the μ_ℓ's are disjoint, then λ is an LCPS on F.

Proposition 3 (Blume et al., 1991) *The following two statements are equivalent.*

1. *(a) $\succsim_\varphi$ satisfies Axioms 1, 2, and 4″ if $\varphi \in 2^F \backslash \{\emptyset\}$, and Axiom 3 if and only if $\varphi \in \Phi$, and (b) the system of conditional preferences $\{\succsim_\varphi | \ \varphi \in \Phi\}$ satisfies Axioms 5 and 6.*
2. *There exist a vNM utility function $\upsilon : \Delta(Z) \to \mathbb{R}$ and a unique LCPS λ on F with support κ such that, $\forall \varphi \in \Phi$,*

$$\mathbf{p}_\varphi \succsim_\varphi \mathbf{q}_\varphi \textit{ iff } \Big(\sum\nolimits_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{p}_\varphi(e))\Big)_{\ell=1}^{L|\varphi} \geq_L \Big(\sum\nolimits_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{q}_\varphi(e))\Big)_{\ell=1}^{L|\varphi},$$

where $\lambda|_\varphi = (\mu'_1, \dots \mu'_{L|\varphi})$ is the conditional of λ on φ (with the LCPS $\lambda|_\varphi$ satisfying, $\forall \ell \in \{1, \dots, L|\varphi\}$, $\mathrm{supp}\mu'_\ell = \pi'_\ell$).

Say that $\succsim_\varphi$ is *conditionally represented* by a vNM utility function υ if (a) $\succsim_\varphi$ is non-trivial and (b) $\mathbf{p}_\varphi \succsim_{\varphi|\{e\}} \mathbf{q}_\varphi$ iff $\upsilon(\mathbf{p}_\varphi(e)) \geq \upsilon(\mathbf{q}_\varphi(e))$ whenever e is deemed subjectively possible. Under Axioms 1′, 2, 3, 4′, 5, and 6 conditional representation follows directly from the vNM theorem of expected utility representation.

Proposition 4 *Assume that (a) $\succsim_\varphi$ satisfies Axioms 1′, 2, and 4′ if $\varphi \in 2^F \backslash \{\emptyset\}$, and Axiom 3 if and only if $\varphi \in \Phi$, and (b) the system of conditional preferences*

$\{\succsim_\varphi |\ \varphi \in \Phi\}$ *satisfies Axioms 5 and 6. Then there exists a vNM utility function* $\upsilon : \Delta(Z) \to \mathbb{R}$ *such that,* $\forall \varphi \in \Phi$, $\mathbf{p}_\varphi \succsim_{\varphi|\{e\}} \mathbf{q}_\varphi$ *iff* $\upsilon(\mathbf{p}_\varphi(e)) \geq \upsilon(\mathbf{q}_\varphi(e))$ *whenever* $e \in \kappa \cap \varphi$.

Under Axioms 1, 2, 3, 4′, 5, 6′, and 16 we obtain the characterization result of Asheim and Perea (2005).

For the statement of this result, we need to introduce the concept of a system of conditional lexicographic probabilities. For this definition, if $\lambda := (\mu_1, \ldots, \mu_L)$ is an LPS and $\ell \in \{1, \ldots, L\}$, then write $\lambda_\ell := (\mu_1, \ldots, \mu_\ell)$ for the LPS that includes only the ℓ top levels of the original sequence of probability distributions.

Definition 1 A *system of conditional lexicographic probabilities* (SCLP) (λ, ℓ) on F with support κ consists of

- An LPS $\lambda = (\mu_1, \ldots, \mu_L)$ on F with support κ
- A function $\ell : \Phi \to \{1, \ldots, L\}$ satisfying *(i)* $\operatorname{supp} \lambda_{\ell(\varphi)} \cap \varphi \neq \emptyset$, *(ii)* $\ell(\varepsilon) \geq \ell(\varphi)$ whenever $\emptyset \neq \varepsilon \subseteq \varphi$, and *(iii)* $\ell(\{e\}) \geq \ell$ whenever $e \in \operatorname{supp} \mu_\ell$

The interpretation is that the conditional belief on φ is given by the conditional on φ of the LPS $\lambda_{\ell(\varphi)}$, $\lambda_{\ell(\varphi)}|_\varphi = (\mu'_1, \ldots \mu'_{\ell(\varphi)|\varphi})$. To determine preference between acts conditional on φ, first calculate expected utilities by means of the top level probability distribution, μ'_1, and then, if necessary, use the lower level probability distributions, $\mu'_2, \ldots, \mu'_{\ell(\varphi)|\varphi}$, lexicographically to resolve ties. The *length function* ℓ thus determines, for every event φ, the number of levels of the original LPS λ that can be used, provided that their supports intersect with φ, to resolve ties between acts conditional on φ.

Condition (i) ensures well-defined conditional beliefs that represent nontrivial conditional preferences. Condition (ii) means that the system of conditional preferences is dynamically consistent, in the sense that strict preference between two acts would always be maintained if new information, ruling out states at which the two acts lead to the same outcomes, became available. To motivate condition (iii), note that if $e \in \operatorname{supp} \mu_\ell$ and $\ell(\{e\}) < \ell$, then it follows from condition (ii) that μ_ℓ could as well ignore e without changing the conditional beliefs.

Proposition 5 (Asheim and Perea, 2005) *The following two statements are equivalent.*

1. *(a)* $\succsim_\varphi$ *satisfies Axioms 1, 2, and 4′ if* $\varphi \in 2^F \backslash \{\emptyset\}$, *and Axiom 3 if and only if* $\varphi \in \Phi$, *and (b) the system of conditional preferences* $\{\succsim_\varphi |\ \varphi \in \Phi\}$ *satisfies Axioms 5, 6′, and 16.*
2. *There exist a vNM utility function* $\upsilon : \Delta(Z) \to \mathbb{R}$ *and an SCLP* (λ, ℓ) *on* F *with support* κ *such that,* $\forall \varphi \in \Phi$,

$$\mathbf{p}_\varphi \succsim_\varphi \mathbf{q}_\varphi \text{ iff } \left(\sum\nolimits_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{p}_\varphi(e))\right)_{\ell=1}^{\ell(\varphi)|\varphi} \geq_L \left(\sum\nolimits_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{q}_\varphi(e))\right)_{\ell=1}^{\ell(\varphi)|\varphi},$$

where $\lambda_{\ell(\varphi)}|_\varphi = (\mu'_1, \ldots \mu'_{\ell(\varphi)|\varphi})$ *is the conditional of* $\lambda_{\ell(\varphi)}$ *on* φ.

By strengthening Axiom 4′ to Axiom 4, we get the following corollary. For the statement of this result, we need to introduce formally the concept of a conditional probability system. A *conditional probability system* (CPS) consists of a collection of subjective probability distributions: If, for each $\varphi \in \Phi$, μ_φ is a subjective probability distribution on φ, and $\{\mu_\varphi \mid \varphi \in \Phi\}$ satisfies $\mu_\varepsilon(\delta) \cdot \mu_\varphi(\varepsilon) = \mu_\varphi(\delta)$ whenever $\delta \subseteq \varepsilon \subseteq \varphi$ and ε, $\varphi \in \Phi$, then $\{\mu_\varphi \mid \varphi \in \Phi\}$ is a CPS on F with support κ.

Corollary 1 *The following three statements are equivalent.*

1. *(a) $\succsim_\varphi$ satisfies Axioms 1, 2, and 4 if $\varphi \in 2^F \backslash \{\emptyset\}$, and Axiom 3 if and only if $\varphi \in \Phi$, and (b) the system of conditional preferences $\{\succsim_\varphi \mid \varphi \in \Phi\}$ satisfies Axioms 5, 6′, and 16.*
2. *There exist a vNM utility function $\upsilon : \Delta(Z) \to \mathbb{R}$ and a unique LCPS $\lambda = (\mu_1, \ldots, \mu_L)$ on F with support κ such that, $\forall \varphi \in \Phi$,*

$$\mathbf{p}_\varphi \succsim_\varphi \mathbf{q}_\varphi \;\textit{ iff }\; \sum_{e \in \varphi} \mu_\varphi(e) \upsilon(\mathbf{p}_\varphi(e)) \geq \sum_{e \in \varphi} \mu_\varphi(e) \upsilon(\mathbf{q}_\varphi(e)),$$

   where μ_φ is the conditional of $\mu_{\ell(\varphi)}$ on φ and $\ell(\varphi) = \min\{\ell \mid \mathrm{supp}\lambda_\ell \cap \varphi \neq \emptyset\}$.
3. *There exist a vNM utility function $\upsilon : \Delta(Z) \to \mathbb{R}$ and a unique CPS $\{\mu_\varphi \mid \varphi \in \Phi\}$ on F with support κ such that, $\forall \varphi \in \Phi$,*

$$\mathbf{p}_\varphi \succsim_\varphi \mathbf{q}_\varphi \;\textit{ iff }\; \sum_{e \in \varphi} \mu_\varphi(e) \upsilon(\mathbf{p}_\varphi(e)) \geq \sum_{e \in \varphi} \mu_\varphi(e) \upsilon(\mathbf{q}_\varphi(e)).$$

4 Concluding Remark

A *full support* SCLP (i.e., an SCLP where $\kappa = F$) combines the structural implication of a full support LPS – namely that conditional preferences are nontrivial – with flexibility w.r.t. whether to assume the behavioral implication of any conditional of such an LPS – namely that the conditional LPS's full support induces preference for cautious behavior. A full support SCLP is a generalization of both

(1) Conditional beliefs described by a single full support LPS $\lambda = (\mu_1, \ldots, \mu_L)$ (cf. Proposition 2): Let, for all $\varphi \in \Phi$, $\ell(\varphi) = L$. With such a maximal length function, the conditional belief on φ is described by the conditional of λ on φ, $\lambda|_\varphi$.

(2) Conditional beliefs described by a CPS (cf. Corollary 1): Let, for all $\varphi \in \Phi$, $\ell(\varphi) = \min\{\ell \mid \mathrm{supp}\lambda_\ell \cap \varphi \neq \emptyset\}$. With such a minimal length function, it follows from conditions (ii) and (iii) of Definition 1 that the full support LPS $\lambda = (\mu_1, \ldots, \mu_L)$ has non-overlapping supports – i.e., λ is an LCPS – and the conditional belief on φ is described by the top level probability distribution of the conditional of λ on φ. This corresponds to the isomorphism between CPS and LCPS noted by Blume et al. (1991, p. 72) and discussed by Halpern (2005) and Hammond (1994).[4]

However, a full support SCLP may describe a system of conditional beliefs that is not covered by these special cases. The following is a simple example: Let $\kappa = F =$

$\{d,e,f\}$ and $\lambda = (\mu_1, \mu_2)$, where $\mu_1(d) = 1/2$, $\mu_1(e) = 1/2$, and $\mu_2(f) = 1$. If $\ell(F) = 1$ and $\ell(\varphi) = 2$ for any other non-empty subset φ, then the resulting SCLP falls outside cases (1) and (2).

As we have seen, a CPS may lead to a refinement of a preference between two acts when new information, ruling out states at which the two acts lead to the same outcomes, becomes available. It might be tempting to ascribe such refinement to the decision maker being initially unaware of the set of states which, according to the new information, contains the true state. However, in the context of sequential equilibrium and rationalizability (which can be modeled by means of a CPS), such an interpretation is inconsistent with the fact that these game-theoretic concepts require each player to consider all information sets of the game, also those which are assigned zero unconditional probability.

It is sometimes claimed that in game-theoretic analysis, LPSs are needed for the analysis of strategic games, while CPSs are suited for extensive games. This paper shows that CPSs are not sufficient for extensive games, for the same reasons that subjective probability distributions do not suffice for strategic games, namely that CPSs cannot handle preference for cautious behavior. In contrast, the concept of SCLP as defined in Definition 1 and characterized in Proposition 5 provides the needed flexibility to account for cautiousness, while ensuring well-defined conditional preferences.

Appendix: Proofs

Proof of Lemma 2. Part 1: Axiom 11 is implied. We must show, under the given premise, that if $e' \gg e''$, then, $\forall f \in F$, $e' \gg f$ or $f \gg e''$. Clearly, $e' \gg e''$ entails $e' \in \kappa$, implying that $e' \gg f$ or $f \gg e''$ if $f \notin \kappa$ or $e'' \notin \kappa$. The case where $f = e'$ or $f = e''$ is trivial. The case where $f \neq e'$, $f \neq e''$, $f \in \kappa$ and $e'' \in \kappa$ remains. Assume that $e' \gg f$ does not hold, which by completeness (Axiom 1) entails the existence of $\mathbf{p}'_{\{e',f\}}$ and $\mathbf{q}'_{\{e',f\}}$ such that $\mathbf{p}'_{\{e',f\}} \precsim_{\{e',f\}} \mathbf{q}'_{\{e',f\}}$ and $\mathbf{p}'_{\{e'\}} \succ_{\{e'\}} \mathbf{q}'_{\{e'\}}$. It suffices to show that $f \gg e''$ is obtained; i.e., $\mathbf{p}_{\{f\}} \succ_{\{f\}} \mathbf{q}_{\{f\}}$ implies $\mathbf{p}_{\{e'',f\}} \succ_{\{e'',f\}} \mathbf{q}_{\{e'',f\}}$. Throughout we invoke Axiom 6 and Lemma 1, and choose $\varphi \in \Phi$ so that $\{e', e'', f\} \in \varphi$.

Let $\mathbf{p}_\varphi \succ_{\varphi|\{f\}} \mathbf{q}_\varphi$. Assume w.l.o.g. that $\mathbf{p}_\varphi(d) = \mathbf{q}_\varphi(d)$ for $d \neq f, e''$, and $\mathbf{p}'_\varphi(d) = \mathbf{q}'_\varphi(d)$ for $d \neq e', f$. By transitivity (Axiom 1), $\mathbf{p}'_\varphi \precsim_{\varphi|\{e',f\}} \mathbf{q}'_\varphi$ and $\mathbf{p}'_\varphi \succ_{\varphi|\{e'\}} \mathbf{q}'_\varphi$ imply $\mathbf{p}'_\varphi \prec_{\varphi|\{f\}} \mathbf{q}'_\varphi$. However, since $\succsim_\varphi$ satisfies Axioms 2 and 4′, $\exists \gamma \in (0,1)$ such that $\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{p}'_\varphi \succ_{\varphi|\{f\}} \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{q}'_\varphi$. Moreover, $\mathbf{p}_\varphi(e') = \mathbf{q}_\varphi(e')$ and $\mathbf{p}'_\varphi \succ_{\varphi|\{e'\}} \mathbf{q}'_\varphi$ entail that $\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{p}'_\varphi \succ_{\varphi|\{e'\}} \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{q}'_\varphi$ by Axiom 2, which implies that $\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{p}'_\varphi \succ_{\varphi|\{e',e''\}} \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{q}'_\varphi$ since $e' \gg e''$. Hence, by transitivity, $\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{p}'_\varphi \succ_{\varphi|\{e',e'',f\}} \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{q}'_\varphi$ – or equivalently, $\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{p}'_\varphi \succ \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{q}'_\varphi$. Now, $\mathbf{q}'_\varphi \succsim_{\varphi|\{e',f\}} \mathbf{p}'_\varphi$ means that $\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{q}'_\varphi \succsim \gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{p}'_\varphi$ by Axiom 2, implying that $\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{q}'_\varphi \succ \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{q}'_\varphi$ by transitivity (Axiom 1), and $\mathbf{p}_\varphi \succ \mathbf{q}_\varphi$ – or equivalently, $\mathbf{p}_\varphi \succ_{\varphi|\{e'',f\}} \mathbf{q}_\varphi$ – by Axiom 2. Thus, $\mathbf{p}_\varphi \succ_{\varphi|\{f\}} \mathbf{q}_\varphi$ implies $\mathbf{p}_\varphi \succ_{\varphi|\{e'',f\}} \mathbf{q}_\varphi$, meaning that $f \gg e''$.

Part 2: Axiom 16 is implied. We must show, under the given premise, that here exists a binary relation $\succsim_F^*$ satisfying Axioms 1, 2, and 4′ such that $\mathbf{p} \succ_{F|\varphi}^* \mathbf{q}$ whenever $\mathbf{p}_\varphi \succ_\varphi \mathbf{q}_\varphi$ and $\emptyset \neq \varphi \subseteq F$. Clearly, since Axiom 6 is satisfied, $\succsim_F$ fulfil these requirements. □

Proof of Proposition 5. 1 implies 2. Since $\succsim_\varphi$ is trivial if $\varphi \notin \Phi$, we may w.l.o.g. assume that Axiom 16 is satisfied with $\succsim_{F|\varphi}^*$ being trivial for any $\varphi \notin \Phi$.

Consider any $e \in \kappa$. Since $\succsim_{\{e\}}$ satisfies Axioms 1, 2, 3, and 4′ (implying Axiom 4 since $\{e\}$ has only one state), it follows from the vNM theorem of expected utility representation that there exists a vNM utility function $\upsilon_{\{e\}} : \Delta(Z) \to \mathbb{R}$ such that $\upsilon_{\{e\}}$ represents $\succsim_{\{e\}}$. By Axiom 5, we may choose a common vNM utility function υ to represent $\succsim_{\{e\}}$ for all $e \in \kappa$. Since Axiom 16 implies, for any $e \in \kappa$, $\succsim_{F|\{e\}}^*$ satisfies Axioms 1, 2, 3, and 4′, and furthermore, $\mathbf{p} \succ_{F|\{e\}}^* \mathbf{q}$ whenever $\mathbf{p}_{\{e\}} \succ_{\{e\}} \mathbf{q}_{\{e\}}$, we obtain that υ represents $\succsim_{F|\{e\}}^*$ for all $e \in \kappa$. It now follows that $\succsim_F^*$ satisfies Axiom 5 of Blume et al. (1991).

By Theorem 3.1 of Blume et al. (1991) $\succsim_F^*$ is represented by υ and an LPS $\lambda = (\mu_1, \dots, \mu_L)$ on F with support κ. Consider any $\varphi \in \Phi$. If $\mathbf{p}_\varphi \succ_\varphi \mathbf{q}_\varphi$ iff $\mathbf{p} \succ_{F|\varphi}^* \mathbf{q}$, then

$$\mathbf{p}_\varphi \succsim_\varphi \mathbf{q}_\varphi \text{ iff}$$
$$\Big(\textstyle\sum_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{p}_\varphi(e))\Big)_{\ell=1}^{L|\varphi} \geq_L \Big(\textstyle\sum_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{q}_\varphi(e))\Big)_{\ell=1}^{L|\varphi},$$

where $\lambda|_\varphi = (\mu'_1, \dots \mu'_{L|\varphi})$ is the conditional of λ on φ, implying that we can set $\ell(\varphi) = L$. Otherwise, let $\ell(\varphi) \in \{0, \dots, L-1\}$ be the maximum ℓ for which it holds that

$$\mathbf{p}_\varphi \succ_\varphi \mathbf{q}_\varphi \text{ if}$$
$$\Big(\textstyle\sum_{e\in\varphi} \mu'_k(e)\upsilon(\mathbf{p}_\varphi(e))\Big)_{k=1}^{\ell|\varphi} >_L \Big(\textstyle\sum_{e\in\varphi} \mu'_k(e)\upsilon(\mathbf{q}_\varphi(e))\Big)_{k=1}^{\ell|\varphi},$$

where the r.h.s. is never satisfied if $\ell < \min\{k\,|\operatorname{supp}\lambda_k \cap \varphi \neq \emptyset\}$, entailing that the implication holds for any such ℓ. Define a set of pairs of acts on φ, $\mathcal{I}$, as follows:

$$(\mathbf{p}_\varphi, \mathbf{q}_\varphi) \in \mathcal{I} \text{ iff}$$
$$\Big(\textstyle\sum_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{p}_\varphi(e))\Big)_{\ell=1}^{\ell(\varphi)|\varphi} =_L \Big(\textstyle\sum_{e\in\varphi} \mu'_\ell(e)\upsilon(\mathbf{q}_\varphi(e))\Big)_{\ell=1}^{\ell(\varphi)|\varphi},$$

with $(\mathbf{p}_\varphi, \mathbf{q}_\varphi) \in \mathcal{I}$ for any acts $\mathbf{p}_\varphi$ and $\mathbf{q}_\varphi$ on φ if $\ell(\varphi) < \min\{\ell\,|\operatorname{supp}\lambda_\ell \cap \varphi \neq \emptyset\}$. Note that $\mathcal{I}$ is a convex set. To show that υ and $\lambda_{\ell(\varphi)}|_\varphi$ represent $\succsim_\varphi$, we must establish that $\mathbf{p}_\varphi \sim_\varphi \mathbf{q}_\varphi$ whenever $(\mathbf{p}_\varphi, \mathbf{q}_\varphi) \in \mathcal{I}$. Hence, suppose there exists $(\mathbf{p}_\varphi, \mathbf{q}_\varphi) \in \mathcal{I}$ such that $\mathbf{p}_\varphi \succ_\varphi \mathbf{q}_\varphi$. It follows from the definition of $\ell(\varphi)$ and the completeness of $\succsim_\varphi$ (Axiom 1) that there exists $(\mathbf{p}'_\varphi, \mathbf{q}'_\varphi) \in \mathcal{I}$ such that

$$\mathbf{p}'_\varphi \succsim_\varphi \mathbf{q}'_\varphi \text{ and } \sum_{e\in\varphi} \mu_{\ell(\varphi)+1}(e)\upsilon(\mathbf{p}'_\varphi(e)) < \sum_{e\in\varphi} \mu_{\ell(\varphi)+1}(e)\upsilon(\mathbf{q}'_\varphi(e)).$$

Objective independence of $\succsim_\varphi$ (Axiom 2) now implies that, if $0 < \gamma < 1$, then

$$\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{p}'_\varphi \succ_\varphi \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{p}'_\varphi \succsim_\varphi \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{q}'_\varphi;$$

hence, by transitivity of $\succsim_\varphi$ (Axiom 1),

$$\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{p}'_\varphi \succ_\varphi \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{q}'_\varphi. \tag{1}$$

However, by choosing γ sufficiently small, we have that

$$\sum_{e\in\varphi} \mu_{\ell(\varphi)+1}(e)\upsilon(\gamma\mathbf{p}_\varphi(e) + (1-\gamma)\mathbf{p}'_\varphi(e))$$
$$< \sum_{e\in\varphi} \mu_{\ell(\varphi)+1}(e)\upsilon(\gamma\mathbf{q}_\varphi(e) + (1-\gamma)\mathbf{q}'_\varphi(e)).$$

Since $\mathcal{I}$ is convex so that $(\gamma\mathbf{p}_\varphi + (1-\gamma)\mathbf{p}'_\varphi, \gamma\mathbf{q}_\varphi + (1-\gamma)\mathbf{q}'_\varphi) \in \mathcal{I}$, this implies that

$$\gamma\mathbf{p} + (1-\gamma)\mathbf{p}' \prec^*_{F|\varphi} \gamma\mathbf{q} + (1-\gamma)\mathbf{q}'. \tag{2}$$

Since (1) and (2) contradict Axiom 16, this shows that $\mathbf{p}_\varphi \sim_\varphi \mathbf{q}_\varphi$ whenever $(\mathbf{p}_\varphi, \mathbf{q}_\varphi) \in \mathcal{I}$. This implies in turn that $\ell(\varphi) \geq \min\{\ell | \operatorname{supp}\lambda_\ell \cap \varphi \neq \emptyset\}$ since $\succsim_\varphi$ is nontrivial. By Axiom 6′, $\ell(\varepsilon) \geq \ell(\varphi)$ whenever $\emptyset \neq \varepsilon \subseteq \varphi$. Finally, since, υ represents $\succsim_{\{e\}}$ for all $e \in \kappa$, it follows that $\mathbf{p}_{\{e\}} \succ_{\{e\}} \mathbf{q}_{\{e\}}$ iff $\mathbf{p} \succ^*_{F|\{e\}} \mathbf{q}$. Hence, we can set $\ell(\{e\}) = L$, implying $\ell(\{e\}) \geq \ell$ whenever $e \in \operatorname{supp}\mu_\ell$.

2 implies 1. This follows from routine arguments. □

Proof of Corollary 1. 1 implies 2. By Proposition 5, the system of conditional preferences is represented by an SCLP (λ, ℓ) on F with support κ. By the strengthening Axiom 4′ to Axiom 4, it follows from the representation result of Anscombe and Aumann (1963) that only the top level probability distribution is needed to represent each conditional preferences; i.e., for any $\varphi \in \Phi$, $\ell(\varphi) = \min\{\ell \mid \operatorname{supp}\lambda_\ell \cap \varphi \neq \emptyset\}$. This implies that any overlapping supports in λ can be removed without changing, for any $\varphi \in \Phi$, the conditional of $\lambda_{\ell(\varphi)}$ on φ, turning λ into an LCPS. Furthermore, the LCPS thus determined is unique.

2 implies 1. This follows from routine arguments.

2 implies 3. $\{\mu_\varphi \mid \varphi \in \Phi\}$ is a CPS on F with support κ since $\mu_\varepsilon(\delta) \cdot \mu_\varphi(\varepsilon) = \mu_\varphi(\delta)$ is satisfied whenever $\delta \subseteq \varepsilon \subseteq \varphi$ and ε, $\varphi \in \Phi$. If an alternative CPS $\{\tilde{\mu}_\varphi \mid \varphi \in \Phi\}$ were to satisfy, for any $\varphi \in \Phi$,

$$\mathbf{p}_\varphi \succsim_\varphi \mathbf{q}_\varphi \text{ iff } \sum_{e\in\varphi} \tilde{\mu}_\varphi(e)\upsilon(\mathbf{p}_\varphi(e)) \geq \sum_{e\in\varphi} \tilde{\mu}_\varphi(e)\upsilon(\mathbf{q}_\varphi(e)),$$

then one could construct an alternative LCPS $\tilde{\lambda} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_L)$ such that, for any $\varphi \in \Phi$, $\tilde{\mu}_\varphi$ is the conditional of $\tilde{\mu}_{\tilde{\ell}(\varphi)}$ on φ, where $\tilde{\ell}(\varphi) := \min\{\ell \mid \operatorname{supp}\tilde{\mu}_\ell \cap \varphi \neq \emptyset\}$, contradicting the uniqueness of λ.

3 implies 2. Construct the LCPS $\lambda = (\mu_1, \ldots, \mu_L)$ by the following algorithm: (i) $\mu_1 = \mu_F$, (ii) $\forall \ell \in \{2, \ldots, L\}$, $\mu_\ell = \mu_\varphi$, where $\varphi = F \backslash \cup_{k=1}^{\ell-1} \mathrm{supp}\mu_k \neq F \backslash \kappa$, and (iii) $\cup_{k=1}^{L} \mathrm{supp}\mu_k = \kappa$. Then, for any $\varphi \in \Phi^a$, μ_φ is the conditional of $\mu_{\ell(\varphi)}$ on φ, where $\ell(\varphi) := \min\{\ell \mid \mathrm{supp}\mu_\ell \cap \varphi \neq \emptyset\}$, and λ is the only LCPS having this property. □

Acknowledgements This paper owes much to Andrés Perea through our cooperation on Asheim and Perea (2005). Moreover, I am grateful to Joe Halpern for clarifying discussions. Finally, I thank participants at LOFT6, FUR XII, and seminar presentations at Université de Montréal, Rochester University and the Norwegian School of Economics and Business Administration for helpful comments.

Notes

[1]This is the terminology introduced by Myerson (1986). In philosophical literature, related concepts are called Popper measures. For an overview over relevant literature and analysis, see Halpern (2005) and Hammond (1994).

[2]One possibility is, following Bewley (1986), to represent incomplete preferences by means of a *set* of subjective probability distributions. See also Aumann (1962).

[3]That is, $\forall \ell \in \{1, \ldots, L|\varphi\}$, $\mu'_\ell = \mu_{k_\ell}|_\varphi$, where the indices k_ℓ are given by $k_0 = 0$, $k_\ell = \min\{k \mid \mu_k(\varphi) > 0 \text{ and } k > k_{\ell-1}\}$ for $\ell > 0$, and $\{k \mid \mu_k(\varphi) > 0 \text{ and } k > k_{L|\varphi}\} = \emptyset$, and where $\mu_{k_\ell}|_\varphi$ is given by the usual definition of conditional probabilities; cf. Definition 4.2 of Blume et al. (1991).

[4]Hence, the isomorphism between CPS and LCPS noted by Blume et al. (1991, p. 72) and discussed by Halpern (2005) and Hammond (1994) implicitly entails that a minimal length function supplements the LCPS. Such a combination is a special case of an SCLP.

References

Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics, 34*, 199–205

Asheim, G. B. (2006). *The consistent preferences approach to deductive reasoning in games*. Dordrecht: Springer

Asheim, G. B., & Dufwenberg, M. (2003). Admissibility and common belief. *Games and Economic Behavior, 42*, 208–234

Asheim, G. B., & Perea, A. (2005). Sequential and quasi-perfect rationalizability in extensive games. *Games and Economic Behavior, 53*, 15–42

Aumann, R. J. (1962). Utility theory without the completeness axiom. *Econometrica, 30*, 445–462

Aumann, R. J. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior, 8*, 6–19

Bewley, T. F. (1986). Knightian decision theory: Part 1. Cowles Foundation DP 807

Blume, L., Brandenburger, A., & Dekel, E. (1991). Lexicographic probabilities and choice under uncertainty, *Econometrica, 59*, 61–79

Halpern, J. Y. (2005). *Lexicographic probability, conditional probability, and nonstandard probability*. Ithaca, NY: Cornell University

Hammond, P. J. (1994). Elementary non-archimedean representations of probability for decision theory and games. In P. Humphreys (Ed.), *Patrick suppes: Scientific philosopher, vol. 1, probability and probabilistic causality* (pp. 25–59). Dordrecht: Kluwer

Kreps, D. M., & Wilson, R. (1982). Sequential equilibria. *Econometrica, 50*, 863–894

Myerson, R. (1986). Multistage games with communication, *Econometrica, 54*, 323–358
Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley
van Damme, E. (1984). A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. *International Journal of Game Theory, 13*, 1–13

Subjective Information in Decision Making and Communication

J.D. Stecher

Keywords: Language · Perceptions · Subjective information · Coordination · Ambiguity

1 Introduction

Fifty years ago, Luce (1956) raised the following example to illustrate problems with neoclassical utility theory:

> Find a subject who prefers a cup of coffee with one cube of sugar to one with five cubes (this should not be difficult). Now prepare 401 cups of coffee with $\left(1+\frac{i}{100}\right)x$ grams of sugar, $i = 0, 1, \ldots, 400$, where x is the weight of one cube of sugar. It is evident that he will be indifferent between cup i and cup $i+1$, for any i, but by choice he is not indifferent between $i = 0$ and $i = 400$.

Luce's purpose was to argue for intransitive indifference. In what follows, I pursue a different approach. Rather than treating this student as having intransitive indifference, with preferences defined over objective cups of coffee, I view the student has having preferences defined over the coffee as he subjectively perceives it. To make this rigorous, I provide an axiomatic model of the student's perceptions, allowing the reader of this article to know which coffee cup the student is tasting, in order to clarify the relationship between the subjective decision problem and the objective alternatives.

The discussion so far looks at the problem in one direction, namely, that a given decision maker can see distinct objects as indistinguishable. The issue in the other direction – that the way a given decision maker perceives a given object may not be unique – is discussed in Husserl (1913), Sect. 41. In describing his view of a table in front of him, Husserl writes

J.D. Stecher
Norwegian School of Economics and Business Administration Helleveien 30, N-5045 Bergen, Norway
e-mail: jack.stecher@nhh.no

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.
© Springer-Verlag Berlin Heidelberg 2008

> I close my eyes. My other senses are inactive in relation to the table. I have now no perception of it. I open my eyes, and the perception returns. The perception? Let us be more accurate. Under no circumstances does it return to me individually the same.

In modern language, Husserl asserts that perception is not a function: a single table, viewed twice but only a blink of the eye apart, need not appear the same way to the same person. Luce's hypothetical student says that the inverse of perception is also not a function. Yet there is some relation between what is objectively present and what the agent perceives. To capture this idea, I model perceptions with two sets, and a binary relation between them. The sets are the objective alternatives (known only to the reader of this paper) and the agent's private, subjective set of alternatives, which I call the agent's *conceptions* of the possible choices. I call the binary relation *perception*. So conception is entirely in the agent's mind, while perception depends on some external stimulus.

There remains the problem of informing a decision maker about a possible choice. When all available alternatives are present, and the decision maker can inspect different alternatives directly, communication about choices may be unproblematic. But there are many situations where decisions are made over long distances: people make purchases in online auctions based on how a seller describes the article for sale. Another example is an investor deciding whether to buy stock in a publicly traded company, where the purchase decision depends on some historical statistics and financial reports, but is unlikely to involve a personal visit to each of the company's factories. To have any sort of economic interaction, different decision makers need to find some set of shared terminology.

Because different decision makers do not necessarily have the same mental conceptions, any shared language necessarily has a certain amount of inherent vagueness: at least one agent is forced to use a language that is an imperfect copy of his conceptions. I capture this in a way that is analogous to the model I present of perceptions: a decision maker's mental conceptions may not have a unique description in the shared language, and the each term in the shared language may not have a unique meaning. For each decision maker, there are two sets of interest for communication: one's subjective mental conceptions (what one wishes to communicate), and the shared terminology; a binary relation links the two.

The model of perception is presented formally in the next section. Section 3 models a decision maker's use of a shared language, and gives necessary conditions on the shared language for avoiding coordination failures. Section 4 concludes.

2 Perception

2.1 Basic Model of Perception

Throughout this paper, fix a set I of individual decision makers. The model of perception consists of two sets and a binary relation between them. The first is the objective set of possible alternatives (real-world objects) labeled X. No decision maker observes X; in fact, a decision maker need not be aware that X exists. The role of X in the model is entirely for the researcher's convenience.

For $i \in I$, there is a set S_i, known only to i, called i's subjective *conceptions*. When no confusion can arise, I drop the subscript and write S. Individual i's preferences are defined on S_i, i.e., on the world as i understands it.

A *perception* is something a decision maker observes when presented with a real-world object. That is, when the stimulus is $x \in X$, the decision maker perceives some subjectively meaningful $a \in S$. A given $x \in X$ need not have a unique conception $a \in S$ as the only way it can be perceived – as in the example from Husserl in Sect. 1 – but a decision maker always has available a coarse enough conception to make every object perceptible. This assumption, though not strictly necessary, simply says that the agent does not make choices among objects that are invisible to him. Conversely, not every conception $a \in S$ is necessarily the perception of a unique $x \in X$, as in Luce's hypothetical student tasting similar cups of coffee.

To formalize this many-to-many relationship, perception is modeled as a binary relation, written as $\Vdash$, between X and S.[1] For $x \in X$ and $a \in S$, the relation $x \Vdash a$ is read as, "the decision maker can perceive real-world object x as subjective conception a." With some technical assumptions, discussed below, the individual's perceptions induce a topology on X, with S as a base of this topology. This means that the individual's preferences, in terms of X, are defined on neighborhoods of possible alternatives.

It is technically convenient in what follows to view the agent's perception relation as a correspondence:

Definition 1. The correspondence $X \xrightarrow{\Vdash} S$ is given by

$$(\forall x \in X) \qquad \Vdash(x) \equiv \{a \in S \,|\, x \Vdash a\}.$$

The inverse correspondence $S \xrightarrow{\Vdash^{-1}} X$ is given by

$$(\forall a \in S) \qquad \Vdash^{-1}(a) \equiv \{x \in X \,|\, x \Vdash a\}.$$

With these definitions, one can define the way an agent perceives some set of alternatives. There are two possible notions of the image of some $D \subseteq X$ along $\Vdash$. The *strong* image gives those conceptions in S that are perceived only when the stimulus is in D. The dual notion is the *weak* image, giving the conceptions in S that can be the perception of some point in D.

Definition 2. The *strong image* of $D \subseteq X$ under $X \xrightarrow{\Vdash} S$ is

$$\Box D \equiv \{a \in S \,|\, \Vdash^{-1}(a) \subseteq D\}.$$

The *weak image* of D under $X \xrightarrow{\Vdash} S$ is

$$\Diamond D \equiv \{a \in S \,|\, \exists x \in \Vdash^{-1}(a) \cap D\}.$$

The notation follows Sambin and Gebellato (1998). By analogy with modal logic, $\Box D$ consists of conceptions that are necessarily the perception of something in D, whereas $\Diamond D$ consists of those conceptions possibly perceived when the stimulus is in D. But the notation is only suggestive – these are not true modal operators, and cannot be iterated.

In an entirely analogous way, the inverse correspondence $S \xrightarrow{\Vdash^{-1}} X$ generates both a strong and weak inverse image of any $U \subseteq S$. In the literature, the strong inverse image is called the *restriction* of U, while the weak inverse image is the *extent* of U; see for example Negri (2002).

Definition 3. The *restriction* of $U \subseteq S$ under $S \xrightarrow{\Vdash} X$ is

$$\text{rest}\, U \equiv \{x \in X \mid \Vdash (x) \subseteq U\}.$$

The *extent* of U under $S \xrightarrow{\Vdash} X$ is

$$\text{ext}\, U \equiv \{x \in X \mid \exists a \in \Vdash (x) \cap U\}.$$

In addition to these four operators, I impose two axioms. First, as indicated above, I require the decision maker to have some way, however coarse, of perceiving any possible alternative:

Axiom 1 *For each $x \in X$, there is some $a \in S$ such that $x \Vdash a$.*

The second axiom, which again is not strictly necessary but technically convenient, says that the decision maker's perceptions are in a sense consistent:

Axiom 2 *For each $a, a' \in S$, if there is some $x \in X$ such that*

$$x \Vdash a \quad \text{and} \quad x \Vdash a',$$

then there is some $a'' \in S$ such that, for all $x' \in X$,

$$x' \Vdash a'' \quad \text{iff} \quad x' \Vdash a \text{ and } x' \Vdash a'.$$

This consistency condition says that a decision maker who can perceive the same object in multiple ways has some conception of objects that can be possibly be perceived in these ways.

2.2 The Topological Interpretation of Perception

This section shows that an individual's perceptions induce a topology on the set of objective choices (i.e., on the real-world objects X). The topological structure is then used in the next section to derive conditions a shared language must have if different individuals can use it without having coordination failures.

The topology is constructed by composing the operations from $X \longrightarrow S$ with the operations $S \longrightarrow X$, yielding operators from $X \longrightarrow X$. I then show that these operators are valid notions of interior and closure.

Intuitively, one can view the members of S as the names of open neighborhoods in the base of a topology on X. That is, each $a \in S$ is associated with

$$\Vdash^{-1}(a) = \{x \in X | x \Vdash a\},$$

which one can view as the neighborhood in X that a names. (Compare Valentini (2001).) Under this interpretation, all of S becomes associated with the family

$$\{\{x \in X | x \Vdash a\} | a \in S\}.$$

Definition 4 (Vickers (1988)). Let X be a topological space with base S. A subset $D \subseteq X$ is *open* iff every $x \in D$ has a neighborhood $N(x) \in S$ such that $N(x) \subseteq D$.

Associating S with neighborhoods in X modifies this definition as follows: $D \subseteq X$ is open if and only if, for every $x \in D$, there is some $a \in S$ such that $x \Vdash a$ and $\Vdash^{-1}(a) \subseteq D$.

Definition 5. The *interior* operator on X induced by $\Vdash$ is

$$\text{int} \equiv \text{ext}\,\Box$$

Thus, for $D \subseteq X$, $\text{int}\,D = \text{ext}\Box D$.

Expanding this definition shows its justification: for $D \subseteq X$,

$$\text{int}\,D \equiv \text{ext}\Box D$$

$$= \{x \in X | (\exists a \in S)(x \Vdash a \text{ and } a \in \Box D)\}$$
$$= \{x \in X | (\exists a \in S)(x \Vdash a \text{ and } (\forall x' \in x)(x' \Vdash a \rightarrow x' \in D))\}.$$

Noting that $x \Vdash a$ iff $x \in \Vdash^{-1}(a)$, the last expression above for $\text{int}\,D$ says that it consists of those points belonging to a neighborhood (named by some $a \in S$) whose points are all in D. That is, the interior of an arbitrary $D \subseteq X$ matches Definition 4.

Conversely, the following holds:

Proposition 1 (Sambin (2001)). *For arbitrary $D \subseteq X$,* int(int D) = int D.

Proof. This will be shown by showing $\Box\text{ext}\Box = \Box$. The result then follows by composing both sides on the left with ext.

Expanding definitions gives, for arbitrary $D \subseteq X$,

$$\Box\text{ext}\Box D = \{a \in S | (\forall x \in X)(x \Vdash a \rightarrow x \in \text{ext}\Box D)\}$$

$$= \{a \in S | (\forall x \in X)(x \Vdash a \rightarrow (\exists a' \in S)(x \Vdash a' \text{ and } (\forall x' \in X)(x' \Vdash a' \rightarrow x' \in D)))\}.$$

The last expression says that $\Box\text{ext}\Box D$ names neighborhoods whose points have a neighborhood a' that is contained in D; thus, $\Box\text{ext}\Box D \subseteq \Box D$.

The reverse inclusion is immediate, by choosing $a' = a$. $\square$

Given that the interior operator matches the classical definition and has the idempotent property in Proposition 1, I make the following definition:

Definition 6. A subset D of X is *open* in the topology induced by $\Vdash$ iff

$$D = \text{ int } D.$$

The standard definition of a closed set is one that contains all of its limit points. That is, the closure of a set is the collection of points for which every open neighborhood intersects the set.

This would suggest that the closure of an arbitrary $D \subseteq X$ should be defined as

$$\{x \in X \,|\, (\forall a \in S)(x \Vdash a \rightarrow (\exists x' \in X)(x' \Vdash a \text{ and } x' \in D))\}.$$

Thus the natural definition of closure in this context is the logical dual of interior:

Definition 7. The *closure* operator on X induced by $\Vdash$ is

$$\text{cl} \equiv \text{ rest}\Diamond$$

Thus, for $D \subseteq X$, cl $D = \text{ rest}\Diamond D$.

An analogous argument to that in the Proposition 1 shows that $\Diamond\text{ext}\Diamond = \Diamond$, and hence that closure is idempotent. This justifies the following definition:

Definition 8. A subset D of X is *closed* in the topology induced by $\Vdash$ iff

$$D = \text{ cl } D.$$

Expanding this definition for $D \subseteq X$ gives

$$\text{cl } D \equiv \text{ rest}\Diamond D$$

$$= \{x \in X \,|\, (\forall a \in S)(x \Vdash a \rightarrow a \in \Diamond D)\}$$
$$= \{x \in X \,|\, (\forall a \in S)(x \Vdash a \rightarrow (\exists x' \in X)(x' \Vdash a \text{ and } x' \in D))\}$$

as desired. Intuitively, a real-world object is in the perceptual closure of D iff every way of perceiving it is a way of perceiving something in D.

It can now be shown that an agent's subjective perceptions induce a topology on X.

Theorem 1 (Perceptual Topology). *The open sets induced by perceptions, under Axioms 1 and 2, form a topology.*

Theorem 1 is proved in a sequence of lemmata:

Lemma 1. *In the perceptual topology, $\varnothing$ is clopen.*

Proof. By definition,

$$\begin{aligned}\text{int } \varnothing = {} & \text{ext}\Box\varnothing \\ = {} & \{x \in X | (\exists a \in S)(x \Vdash a \text{ and } (\forall x' \in X)(x' \Vdash a \to x' \in \varnothing))\} = \varnothing.\end{aligned}$$

Thus, $\varnothing$ is open.

Analogously,

$$\begin{aligned}\text{cl } \varnothing = {} & \text{rest}\Diamond\varnothing \\ = {} & \{x \in X | (\forall a \in S)(x \Vdash a \to (\exists x' \in X)(x' \Vdash a \text{ and } x' \in \varnothing))\} = \varnothing.\end{aligned}$$

Thus, $\varnothing$ is closed. □

Lemma 2. *In the perceptual topology, X is clopen.*

Proof. By definition,

$$\begin{aligned}\text{int } X = {} & \text{ext}\Box X \\ = {} & \{x \in X | (\exists a \in S)(x \Vdash a \text{ and } (\forall x' \in X)(x' \Vdash a \to x' \in X))\} \\ = {} & \{x \in X | (\exists a \in S) x \Vdash a\}.\end{aligned}$$

By Axiom 1, this is all of X, so X is open.

Analogously,

$$\begin{aligned}\text{cl } X = {} & \text{rest}\Diamond X \\ = {} & \{x \in X | (\forall a \in S)(x \Vdash a \to (\exists x' \in X)(x' \Vdash a \text{ and } x' \in X))\} = X.\end{aligned}$$

Thus, X is closed. □

Lemma 3. *The union of open sets in the perceptual topology is open.*

Proof. By definition,

$$\begin{aligned}\bigcup_\alpha \text{int } D_\alpha = {} & \{x \in X | (\exists \alpha)(\exists a \in S)(x \Vdash a \text{ and } (\forall x' \in X)(x' \Vdash a \to x' \in D_\alpha))\} \\ \subseteq {} & \{x \in X | (\exists a \in S)(x \Vdash a \text{ and } (\forall x' \in X)(x' \Vdash a \to (\exists \alpha)(x' \in D_\alpha)))\} \\ = {} & \text{int} \bigcup_\alpha \text{int } D_\alpha.\end{aligned}$$

To see the reverse inclusion, note that for any $D \subseteq X$, if $x \in \text{ int } D$, then

$$(\exists a \in S)(x \Vdash a \text{ and } (\forall x' \in S)(x' \Vdash a \to x' \in D)).$$

Picking $x' = x$ gives

$$x \in \text{ int } D \to x \in D,$$

i.e., $\text{int } D \subseteq D$. In particular,

$$\text{int} \bigcup_\alpha \text{int } D_\alpha \subseteq \bigcup_\alpha \text{int } D_\alpha.$$

Combining these gives the result. □

Lemma 4. *The intersection of finitely many open sets in the perceptual topology is open.*

Proof. It suffices to show that the intersection of two open sets is open, as the result then follows by induction. For $D, E \subseteq X$,

$$\text{int}\, D \bigcap \text{int}\, E = \{x \in X | (\exists a, b \in S)(x \Vdash a \text{ and } x \Vdash b$$

$$\text{and } (\forall x' \in X)(x' \Vdash a \to x' \in D) \text{ and } (\forall x'' \in X)(x'' \Vdash b \to x'' \in E))\}.$$

By Axiom 2, if $x \Vdash a$ and $x \Vdash b$, then there is some $c \in S$ such that $x \Vdash c$ and

$$(\forall x' \in X)(x' \Vdash c \to x' \Vdash a \text{ and } x' \Vdash b),$$

which in turn implies,

$$(\forall x' \in X)(x' \Vdash c \to x' \in D \bigcap E).$$

Thus, $\text{int}\, D \bigcap \text{int}\, E \subseteq \text{int}\, (D \bigcap E)$.

Conversely, if $x \in \text{int}\, (D \bigcap E)$, then there is a neighborhood $a \in S$ of x that is contained in $D \bigcap E$, which means that $\Vdash^{-1}(a) \subseteq D$ and $\Vdash^{-1}(a) \subseteq E$. This says that $x \in \text{int}\, D \bigcap \text{int}\, E$, which by the arbitrariness of x implies $\text{int}(D \bigcap E) \subseteq \text{int}\, D \bigcap \text{int}\, E$.

Combining these shows that the finite intersection property holds. □

3 Communication

3.1 Basic Reporting Model and Topological Interpretation

For two agents to communicate, and in particular to use a language for trade, they must be able to reach some sort of understanding about what they are offering or requesting in exchange. Trades cannot be stated in terms of X, as agents do not observe X directly, or even have mental conceptions of what is in X. Moreover, an agent cannot offer directly some object as perceived in S_i: the nature of mental conceptions is that they are private and subjective. A shared language offers a way around this difficulty. The terminology in the shared language is public, so that the agents can use the language to try to validate (by consensus) which objects are under discussion.

A language is modeled as a set T of shared terminology, and, for each $i \in I$, a binary relation R_i between individual i's private conceptions S_i and T. The relation represents i's private semantics, i.e., how i can honestly describe some $a \in S_i$. For an illustration of how the language can evolve, see Ahn (2000).

A term $t \in T$ may be a valid report for more than one $a \in S_i$. Conversely, the same conception may have more than one way it can be reported. Luce's hypothetical student in Sect. 1 may thus have many cups of coffee he describes as medium sweet; conversely, many cups of coffee – even some that the student might be able to distinguish – could still be ones he might characterize as medium sweet.

For $(a,t) \in S_i \times T$, $aR_i t$ is read as, "a can be reported as t by agent i." In parallel with the discussion on perceptions, there are two correspondences associated with the agent's reporting relation:

$$(\forall a \in S_i) \qquad R_i(a) \equiv \{t \in T | aR_i t\}$$

and

$$(\forall t \in S_i) \qquad R_i^{-1}(t) \equiv \{a \in S_i | aR_i t\}.$$

The inverse correspondence gives the agent's interpretation of what a report means. These two correspondences generate the operations $\Diamond$, ext, $\Box$, and rest. Thus, for $U \subseteq S_i$ and $W \subseteq T$,

$$\Diamond_{R_i}(U) \equiv \{t \in T | \exists a \in R_i^{-1}(t) \cap U\},$$

$$\Box_{R_i}(U) \equiv \{t \in T | R_i^{-1}(t) \subseteq U\},$$

$$\text{ext}_{R_i}(W) \equiv \{a \in S_i | \exists t \in R_i(a) \cap W\},$$

and

$$\text{rest}_{R_i}(W) \equiv \{a \in S_i | R_i(a) \subseteq W\}.$$

Reporting can be interpreted in terms of the topology (called the *reporting topology*) on an agent's conceptions. As in Theorem 1, I impose two axioms:

Axiom 3 *For each $a \in S_i$, there is a $t \in T$ such that $aR_i t$.*

Axiom 4 *For each $t, t'' \in T$, if there is an $a \in S_i$ such that*

$$aR_i t \quad \text{and} \quad aR_i t',$$

then there is some $t'' \in T$ such that, for all $a' \in S_i$,

$$a' R_i t'' \quad \text{iff} \quad a' R_i t \quad \text{and} \quad a' R_i t'.$$

Axioms 3 is a non-degeneracy requirement. It says that there must be some way, however vague, of reporting anything the agent may want to report. That is, the language must have some sufficiently broad terms ("stuff," for example) to cover anything. Axiom 4 requires consistency of the language: if there are conceptions that can be reported more than one way, there must be a way to express that there are multiple possible reports. Thus, if a cup of coffee could accurately be described as medium sweet or as lightly sweetened, then there must be a way to say that the coffee is light-to-medium sweetness. As in Theorem 1, this makes it easier to guarantee that the induced topology satisfies the finite intersection property.[2]

The following definitions are analogous to those under perception:

Definition 9. The *reporting interior* operator is

$$\text{int}_{R_i} \equiv \text{ext}_{R_i} \Box_{R_i}.$$

The *reporting closure* operator is

$$\mathrm{cl}_{R_i} \equiv \ \mathrm{rest}_{R_i} \Diamond_{R_i}.$$

A subset U of S_i is *open* in the reporting topology if and only if $U = \mathrm{int}_{R_i}(U)$, and is *closed* in the reporting topology if and only if $U = \mathrm{cl}_{R_i}(U)$.

Theorem 2 (Reporting Topology). *The open sets induced by the agent's reporting relation, under Axioms 3 and 4, form a topology.*

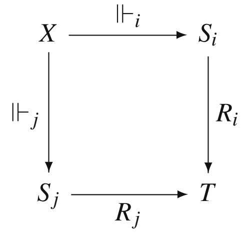
Proof. Entirely analogous to the proof of Theorem 1. □

3.2 Communication

The topological interpretations of perceptions and semantics make it possible to address how agents with fundamentally different worldviews can nevertheless find a reliable way to trade. What is essential is that something one agent offers in the shared language must be something that a trading partner subjectively interprets as a faithful representation of what the trading partner eventually perceives. Because this subjective notion of a faithful representation differs from the accountant's objective notion FASB (1980), I refer to this idea as *heterogeneous faithfulness*.

As neither reports nor perceptions are in general unique, the shared language cannot guarantee that the same agent *necessarily* issues the same report when faced with the same object. The most that can be required is that one agent reports what he or she sees in a way that a trading partner would agree is a valid possible report.

Definition 10. Let $i, j \in I$ be two agents, with conceptions S_i, S_j, perception relations $\Vdash_i, \Vdash_j$, and reporting relations R_i, R_j for a set of common terminology T. The language is *heterogeneously faithful* between i and j if and only if the following diagram commutes:



That is, the language is heterogeneously faithful if and only if

$$R_i \circ \Vdash_i = R_j \circ \Vdash_j \ .$$

If this holds for every $i, j \in I$, then the language is said to be *heterogeneously faithful*.

The following proposition shows that the direction of the definition could be reversed; that is, an equivalent requirement is that two agents have compatible interpretations of the language.

Proposition 2. *Suppose a language is heterogeneously faithful between two agents, $i, j \in I$. Then the interpretation of the reports is also heterogeneously faithful, i.e.,*

$$\Vdash_i^{-1} \circ R_i^{-1} = \Vdash_j^{-1} \circ R_j^{-1}.$$

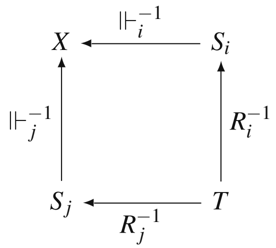
Proof. The heterogeneous faithfulness between i and j means that, given $x \in X$ and $t \in T$, agent i can report x as t iff agent j can do so also. For i to be able to report x as t, there must be some $a \in S_i$ that is a way i can perceive x which i can report as t:

$$(\exists a \in S_i)(x \Vdash_i a \text{ and } a R_i t),$$

which also says that i can interpret t as x, i.e., that

$$(\exists a \in S_i)(a \in R_i^{-1}(t) \text{ and } x \in \Vdash_i^{-1}(a)).$$

By an identical argument, if j can report x as t, then j can interpret t as x. Thus, heterogeneous faithfulness between i and j means that the following square also commutes:



□

Proposition 2 says that a language is useful for speakers if and only if it is useful for listeners. This feature depends on the fact that reporting and perception are defined by binary relations and not necessarily by functions.

Theorems 1 and 2 show that $\langle X, S_i \rangle$ and $\langle S_j, T \rangle$ are topological spaces. Heterogeneous faithfulness thus requires each agent's perceptions to induce a correspondence that carries collections of open neighborhoods to collections of open neighborhoods, i.e., that takes open sets to open sets. Thus, heterogeneous faithfulness is essentially lower hemi-continuity (Berge, 1963):

Definition 11. A correspondence $X \xrightarrow{\Vdash} S$ is *lower hemi-continuous* if, for every $U \subseteq S$,

$$\{x \in X | \exists a \in \Vdash (x) \cap \text{ int } U\}$$

is open.

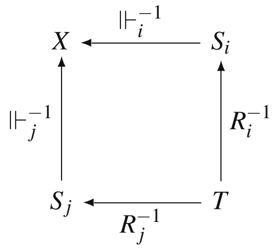
Theorem 3. *If a language is heterogeneously faithful between two agents $i, j \in I$, then j's perception correspondence $\Vdash_j (\cdot)$ is a lower hemi-continuous correspondence between X, endowed with i's perceptual topology, and S_j, endowed with j's reporting topology.*

Conversely, suppose $\Vdash_j(\cdot)$ *is a lower hemi-continuous correspondence between the topological spaces* $\langle X, S_i\rangle$ *and* $\langle S_j, T\rangle$*. Then, nonconstructively, there exists a reporting relation* $R_i(\cdot)$ *for i that makes the language heterogeneously faithful.*

Proof. By the proof of Proposition 1, $\Box\,\text{ext}\,\Box = \Box$, and an analogous argument shows that $\text{ext}\,\Box\,\text{ext} = \text{ext}$.

If $U \subseteq S_j$ is the extent of some $W \subseteq T$, then $\text{int}\,U = \text{int}(\text{ext}(W)) = \text{ext}\,\Box\,\text{ext}\,W = \text{ext}\,W = U$. Conversely, if $U = \text{int}\,U$, then automatically U is the extent of some $W \subseteq T$, namely $\Box U$. Thus $U \subseteq S_j$ is open in the reporting topology if and only if it is the extent of some $W \subseteq T$, i.e., iff it is the inverse image of some subset of T along R_j. A similar argument holds for an open $D \subseteq X$ in i's perceptual topology.

The definition of lower hemi-continuity thus says that the inverse image of any $W \subseteq T$ along $\Vdash_j^{-1} \circ R_j^{-1}$ is the extent of some subset $U' \subset S_i$. So if $W = \text{ext}_{R_i} U'$ for some $U' \subseteq S_i$, then the relation $\Vdash_j$ is lower hemi-continuous. But this just says that the square below commutes:



By Proposition 2, this is equivalent to heterogeneous faithfulness. Therefore, heterogeneous faithfulness implies lower hemi-continuity.

For the second part of the theorem, define the reporting relation for i by $R_i \equiv R_j \circ \Vdash_j \circ \Vdash_i^{-1}$. The continuity of $\Vdash_j$ means that R_i takes open neighborhoods in i's perceptual topology to open neighborhoods in j's reporting topology, which is just the definition of heterogeneous faithfulness. $\Box$

Remark 1. The non-constructivity of the second part of Theorem 3 means that the way agents form a heterogeneously faithful language remains an open problem. To construct the desired reporting relation for agent i, one would need access to S_i and S_j (along with the various relations that are composed). This would imply knowledge of others' perceptions and of X, but the phenomena being studied is that no one has such knowledge.

Accordingly, what the latter part of Theorem 3 establishes is that the non-existence of the desired relation is contradictory. To one who is omniscient, this is equivalent to existence, but it is clear that the relation used in the proof could not in practice be constructed.

There are, however, some examples of languages that are heterogeneously faithful and could be constructed. A trivial case is the universal language, where T is a singleton (e.g, {"stuff"}). In this case, the reporting relations are trivially faithful.

This language is not especially helpful, as any agents who agree to trade have no information on what they are bargaining for, and provide no information on what they are offering; thus, any economy that uses this language is a grab bag.

4 Conclusion

This paper studies the decision-theoretic consequences of subjective information, by modeling how an agent perceives real-world objects, and by modeling how such an agent can nevertheless use a shared language. The framework here is then used to study the interaction between subjective perceptions and subjective semantics, and how these interact to put restrictions on any language that can be used without coordination failures.

The models of perceptions and of use of a language have useful topological interpretations. The connection between what an agent perceives and what an agent reports is shown to be a form of continuity in these topological spaces. Intuitively, the continuity condition says that an agent reports a real-world object in a way that a trading partner, viewing the same object, could see as justified. The practical implication is that if a language enables communication that is overly precise, one agent can bargain for something distinct to him but indistinguishable to his trading partner.

The notion of agents' beliefs thus arises in a different sense from the usual probabilistic one. One agent may observe another agent's use of a language, and infer the distinctions that the other agent is capable of making. Thus an extension of the current model to a dynamic model would enable one to discuss beliefs about another agent's subjective understanding of the world.

Acknowledgement This paper is based on Chaps. 2 and 4 of my Ph.D. dissertation, *Business Language and Asymmetric Perceptions*, defended in 2005. Thanks are due to my advisor, John Dickhaut, and to committee members John Kareken, Regina Anctil, Beth Allen, and Cris Calude. I also received useful comments from René Ahn, Douglas Bridges, Ed Green, Chandra Kanodia, Sara Negri, Jan von Plato, Kira Pronin, Marcel Richter, Giovanni Sambin, Haiwen Zhang, and participants at various conferences (Venice 2003 on constructive mathematics, SABE 2003, Econometric Society 2004 Australasian meeting, Association of Symbolic Logic 2005 annual meeting, TARK X, and the FUR XII). All errors are my own.

Notes

[1]This symbol is the forcing relation introduced by Cohen. For background and historical discussion, see Avigard (2004).

[2]For an approach not requiring these axioms, see Walicki, Wolter, and Stecher (2006).

References

Ahn, R. M. C. (2000). *Agents, objects, and events: A computational approach to knowledge, observation, and communication*. Unpublished doctoral dissertation, Technische Universitiet Eindhoven

Avigard, J. (2004). Forcing in proof theory, *Bulletin of Symbolic Logic, 10*(3), 305–33

Berge, C. (1963). *Topological spaces*. Oliver and Boyd

FASB. (1980). *Qualitative characteristics of accounting information*. (Statement of Financial Accounting Concepts No. 2). Financial Accounting Standards Board

Husserl, E. (1913). *Ideas*. First Collier Books. (Translated from the German, *Ideen au einer reinen Phämoneologie und phänomenologischen Philosophie*. English edition first published 1931)

Luce, R. D. (1956). Semiorders and a theory of utility discrimination, *Econometrica, 24*(2), 178–91

Negri, S. (2002). Continuous domains as formal spaces, *Mathematical Structures in Computer Science, 12*, 19–52

Sambin, G. (2001). *The basic picture, a structure for topology (the basic picture, I)*. (Tech. Rep.). Italy: University of Padua, Department of Pure and Applied Mathematics

Sambin, G., & Gebellato, S. (1998). A preview of the basic picture: A new perspective on formal topology. In *Types and proofs for programs* (Vol. 1657, pp. 194–207) Berlin: Springer

Valentini, S. (2001). *Fixed points of continuous functions between formal spaces* (Tech. Rep.). Italy: University of Padua, Department of Pure and Applied Mathematics

Vickers, S. J. (1988). *Topology via logic*. (No. 5). Cambridge: Cambridge University Press

Walicki, M., Wolter, U., & Stecher, J. (2006). The choice of standards for a reporting language. In S. S. Goncharov, H. Ono, & R. G. Downey (Eds.), *Mathematical logic in Asia: Proceedings of the ninth Asian logic conference*. Singapore: World Scientific

Part II
Risk Modeling

Sensitivity Analysis in Decision Making: A Consistent Approach

E. Borgonovo(✉) and L. Peccati

Keywords: Local and global sensitivity analysis · Constrained sensitivity analysis · Importance measures · von Neumann–Morgenstern expected utility · Bayesian decision theory

1 Introduction

Sensitivity analysis is "*an integral part* (Celemen, 1997)" of any decision-making process accompanied by the creation of a decision-support model (see also Saltelli, Tarantola, and Campolongo (2000)).

Several works (Insua, Ruggeri, & Martin, 1998; Wallace, 2000) show that the use of sensitivity analysis in decision theoretic problems requires special attention and shares distinctive features with respect to the sensitivity analysis of generic simulation models. This becomes especially relevant when the decision-maker interrogates the decision-support model to further manage the problem or to gain additional insights during the preferred alternative implementation phase. In fact, as a decision-support model is (or should be) developed in accordance with the underlying theory, in the same way the sensitivity analysis method ought to conform to such theory. Techniques for the problem of stability of a preferred alternative to *imprecision* (Fishburn, Murhpy, & Isaacs, 1968) in the probabilities are discussed in Evans (1984) and generalized in Ringuest (1997). Bushena and Zilberman (1999) utilize sensitivity analysis to test von Neumann–Morgenstern (vNM) expected utility violations. Pauwels, Van De Walle, Hardeman, and Soudan (2000) utilize sensitivity analysis to assess the robustness of their model results to changes in "*the various model parameters*". In Bayesian statistics, we recall the seminal works of

E. Borgonovo
IMQ, Bocconi University, Viale Isonzo 25, 20135, Milano, Italy
e-mail: emanuele.borgonovo@unibocconi.it

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*.
Theory and Decision Library C 42.
© Springer-Verlag Berlin Heidelberg 2008

Berger (1984, 1990), Gustafson (1996) and Gustafson and Wasserman (1995) that deal with the sensitivity of a posterior (distribution and expectation) to the prior.[1]

Samuelson (1947) defines sensitivity analysis as the problem of addressing "*the response of our system to changes in certain parameters*". At the Sensitivity Analysis of Model Output conference in 2000, a more recent definition of sensitivity analysis sounds: "*the study of how the variation in the output of a model can be apportioned to variations in the input* (Tarantola, 2000)". We note that this second definition can be equivalently expressed as the determination of the influence or importance of an input (Saltelli, 2002).

It is the purpose of this work to build the framework necessary to respond to Samuelson's and Tarantola–Saltelli's definitions in a rigorous fashion. We state first the elements of the problem in the context of decision-support models. Samuelson's *system* is the decision-making problem at hand. The *parameters* are those exogenous variables that the decision-maker needs to assign in order to obtain a quantitative answer from the model. The output is the value of the decision-support criterion estimated by the model. The system response is the change in decision-support criterion, which follows from a different selection of the parameter values. Then, the two definitions of sensitivity analysis of Samuelson (1947) and Tarantola (2000)/Saltelli (2002) turn into the following two questions:

Question 1 *The determination of the response of the decision-support criterion to changes in the parameters;*

and

Question 2 *The study of how the variation in the decision-support model output can be apportioned to variations in the parameters.*

We define and compare the sensitivity analysis problems faced by decision-makers with two different states of belief in the parameter values. First, a decision-maker who is provided with a certain (objective) value of the parameters. We refer to this decision-maker as to a vNM decision-maker see Smith and von Winterfeld (2004, p. 562). In fact, if a decision-maker possesses objective values of the parameters, She/He is also equipped with objective values of the probabilities – in building a decision-support model, probabilities are a subset of the parameters, in general. – Therefore, the decision-maker needs to select alternatives developing the decision-support model consistently with the vNM axioms (von Neumann and Morgenstern, 1947). Then, we consider the sensitivity analysis problem from the perspective of a decision-maker who is not capable of fixing the probabilities at a certain value, but whose state of belief allows Her/Him to assign prior distributions. We refer to this decision-maker as to a Bayesian decision-maker (see Smith and von Winterfeld (2004, p. 563). In this case, the decision-maker develops the model according to the axioms of Pratt, Raiffa, and Schlaifer (1995).

We show that the answers to the two questions define completely different mathematical frameworks for the two decision-makers. For a vNM decision-maker one has a local sensitivity analysis problem, as a vNM the decision-maker is equipped with an objective value of the probabilities, (say p). The answer to Questions 1 and 2 must be studied "at p". In this case, the answer to Question 1 is found by applying comparative statics (Samuelson, 1947) the answer to Question 2 is found through the differential importance measure (Borgonovo, 2008). However, if the decision-maker were to utilize these approaches as such, She/He could draw misleading conclusions. In fact, most of sensitivity analysis methods are built under the assumption that parameters (in our case the probabilities) are free to vary. However, probabilities feeding into a decision-support model are constrained, as their values must sum to unity. Thus, to perform the sensitivity analysis of a vNM model, one needs to extend comparative statics and the differential importance to the case of constrained inputs. To do so, we make use of a recent result (Borgonovo, 2006b) introducing the concept of constrained derivative. The result allows us to answer Questions 1 and 2 and derive the analytical expressions both of the comparative statics and the importance of probabilities in vNM models.

We then consider a decision-maker who decides to gather data on inputs, assigns prior distributions, and updates Her/His distributions as new evidence becomes available (see Pratt et al. (1995)). The decision-maker is now a Bayesian one, and we assume that She/He also develops the model consistently with the subjective expected utility framework. As an immediate consequence, the techniques used to answer Questions 1 and 2 for a vNM decision-maker reveal themselves unsatisfactory, as the analysis cannot be performed "at p". To deal with the problem we introduce the input parameter measure space $(\Theta \subseteq \mathbb{R}^n, B(\Theta), \mu)$ (see Gustafson (1996)) and let $\pi(\cdot)$ the density with respect to μ reflecting the decision-maker state of belief in the inputs. Question 1 is now answered by studying the sensitivity of the expected utility to changes in $\pi(\cdot)$, i.e., the response of the model is no more studied "at p". To answer to Question 1 we make use of the concept of perturbation introduced in Gustafson (1996), and, by means of Fréchet differentiation, we derive the expression of the rate of change in the expected utility in the direction of the perturbation. We show that to answer Question 2 in a Bayesian setting, no change in prior/posterior is needed but the analysis must be performed with reference to the current decision-maker state of belief. We consider the set of all conditional utilities $U_{|\theta_s}$, given that θ_s assumes a certain value. We show that the relevance of θ_s ($s = 1, 2, \ldots, n$) on U given the current decision-maker state of belief (i.e., without any change in prior or posterior) can be found using:

$$\delta_s = \frac{1}{2}\mathbb{E}_\mu \left[d(U, U_{|\theta_s}) \right] \tag{1}$$

with $d(U, U_{|\theta_s})$ the distance between the unconditional and conditional utility (see Borgonovo (2006a, 2007)). We also compare the meaning of δ_s to the expected value of sample information on θ_s ($EVSI_s$).

The remainder of the paper is organized as follows. Section 2.1 presents the mathematical framework of constrained sensitivity analysis necessary to perform the sensitivity analysis of vNM models. Section 2.2 derives analytical results for the answers to Questions 1 and 2 in vNM models. Section 3 deals with the conditions that allow to extend the results obtained for a vNM decision maker to the case of uncertainty in the parameters. Section 4 deals with the formulation of the sensitivity analysis questions for a Bayesian decision maker. Section 5 offers a summary of the findings and future research perspectives.

2 Sensitivity Analysis for Decision Analysis Models with Given Probabilities

2.1 Local Constrained Sensitivity Measures

In this section, starting with a decision analysis example, we show general results that are going to be utilized to derive sensitivity measures for answering Questions 1 and 2 in vNM models.

In most of the applications, the initial modeling effort and time is devoted to the identification of random events and alternatives involved in the problem. Once a credible decision-support model has been built, analysts start gathering information on numbers. The first model runs are usually performed with the inputs fixed at the so called base case (reference) value (let us call it p). If one were to stop the analysis at this moment, one would be stating that the decision-maker's state of belief on the inputs is reflected by a precise point of the input parameter space. Many works refer to this fact saying that that decision-makers are not willing/able to assess probability distributions or that "*practical arguments against probability distributions*" emerge (Wallace, 2000). Van Groenendaal (1998) proposes to perform sensitivity analysis without assessing probability distributions, since, if these distributions cannot be precisely estimated, one would be better off not estimating them at all. However, as (Wallace, 2000) recognizes, such arguments are of a practical nature, and not of a theoretical one.[2]

In His classical work, Howard (1988) emphasizes the role of sensitivity analysis in decision-making and introduces Tornado Diagrams (see also Eschenbach (1992)). In the well known textbook of Celemen (1997), sensitivity analysis explicitly regarded as one of the main steps of the decision analysis process (Fig. 1, Chap. 1 of Celemen (1997)).

Standard decision analysis software enables the performance of sensitivity analysis mainly through a one parameter at a time approach, and subroutines displaying Tornado diagrams or break-even analysis are available. The way such diagrams are obtained is by fixing all the parameters, but the one of interest, at their reference value and registering in a single shot (Tornado diagram) or step by step (break-even analysis) the change in the value of the preferred alternative (Tornado Diagram) or of all alternatives (break-even analysis). In addition, Tornado Diagrams are utilized to "*answer the question: what matters in this decision*? (Celemen, 1997)," i.e., to answer Question 2.

We show by means of an example that, if one were to apply Tornado diagrams to a model with non binary events, one would not be able to carry out a sensitivity analysis while respecting the definition of the diagrams themselves.

Example 1. Consider the oil wildcatter problem as reported in Smith (1998).

1. The decision-maker is interested in the sensitivity of the problem w.r.t. the probabilities. Suppose that He/She wishes to apply one of the sensitivity analysis schemes discussed in Chap. 5 of Celemen (1997), a Tornado diagram, say, letting

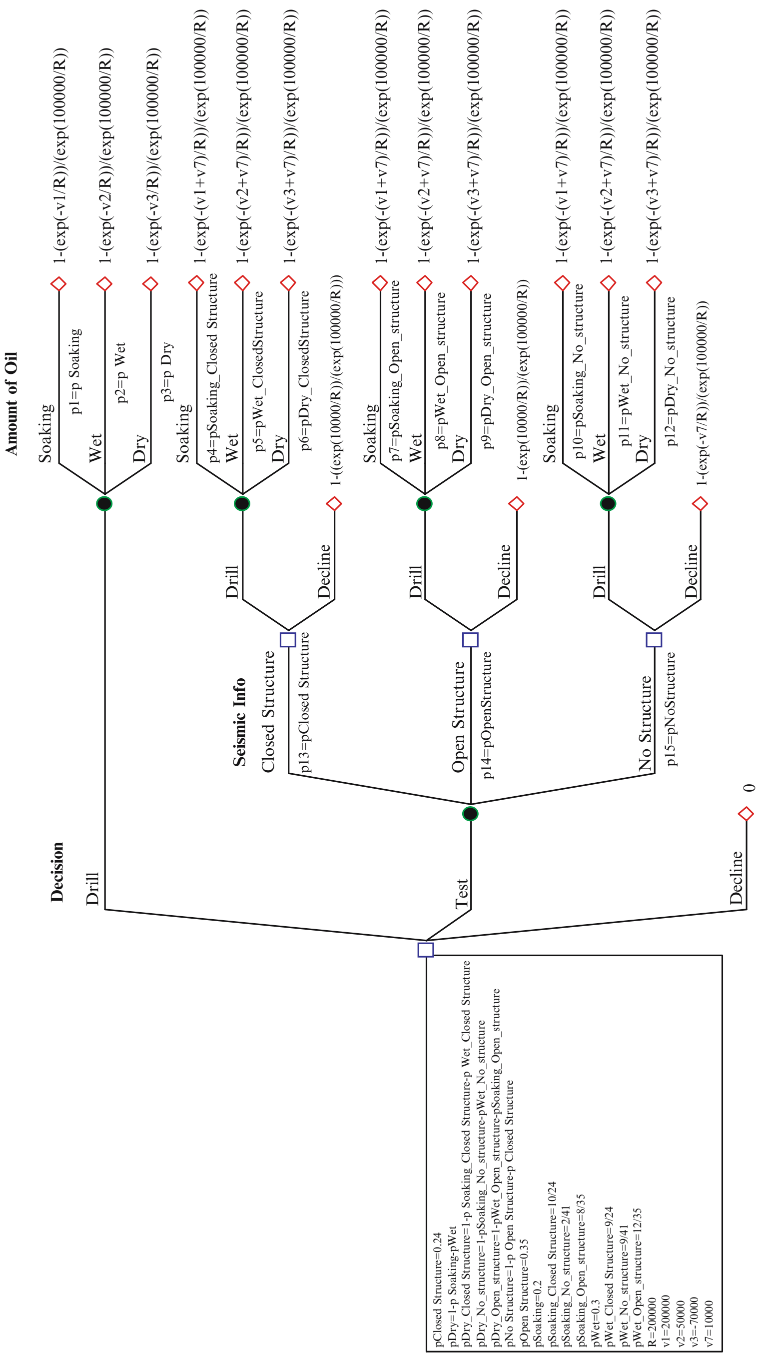


Fig. 1 Decision Tree representation of the oil-wildcatter problem: the presence of non-binary events reveals the constrained nature of the sensitivity analysis on the probabilities

$p_{Soaking}$ vary from $p^0_{Soaking} = 0.2$ to $p^\Delta_{Soaking} = 0.3$. Strictly applying the definition of a Tornado diagram, one would have to keep p_{Wet} at 0.3 and p_{Dry} at 0.5. But in so doing one would violate the laws of probability since $p_{Wet} + p_{Soaking} + p_{Dry} \neq 1$.[3] To solve the problem, one can try the following. Rewrite the constraint defining p_{Wet} as a function of p_{Dry} and $p^*_{Soaking}$: $p_{Wet} = 1 - p_{Dry} - p^\Delta_{Soaking}$. Substituting for $p^\Delta_{Soaking} = 0.3$ and $p_{Dry} = 0.5$, one finds $p_{Wet} = 0.2$.

2. Suppose now that the decision-maker is interested in letting $p^\Delta_{Soaking} = 0.6$. As in the above example, He/She solves for p_{Wet} to get: $p_{Wet} = 1 - p_{Dry} - p^\Delta_{Soaking} = -0.1$[4] which violates another law of probability, since no negative probabilities are allowed. If, instead, the decision-maker solves for p_{Dry}, He/She would get the answer: $p_{Dry} = 1 - p_{Wet} - p^\Delta_{Soaking} = 1 - 0.3 - 0.6 = 0.1$ which allows the sensitivity to be performed.

Part 1 of Example 1 shows that the sensitivity on one of the probabilities ($p_{Soaking}$ in our case) provokes a change in the other probabilities (in this case p_{Wet} moves from 0.5 to 0.4). Numerically it is due to the fact that the probabilities must sum to unity. Conceptually the reason lies in the fact that the measure (or probability distribution) of all events related to the same node in the diagram changes, of the measures of one of the events changes.

These facts make the problem of the sensitivity of decision-support models different from the classical Sensitivity analysis framework of Samuelson (1947) and from the more recent framework of Tarantola (2000). In fact, such definitions consider the parameters as “free” to vary. However, as our example underlines, there are models for which parameters variations are linked by external relationships. We use the term constrained sensitivity. Part 2 of the example shows that when parameters are constrained, then the sensitivity depends on the way the constrained is solved. More precisely, it depends on which of the parameters is selected as a dependent one. In fact, in practical applications, one utilizes the probability of the last outcome as dependent one (see, for example, Sect. 5.3 of Celemen (1997)), but there is no reason while the first one could be selected as dependent. The above facts find their explanation in the following Theorem (the proof can be found in Borgonovo (2006b)).

Theorem 1. *Let*

$$Y = f(x),\ x \in A \subseteq \mathbb{R}^n \to \mathbb{R} \tag{2}$$

Let x be partitioned in $Q < n$ groups of parameters x^q ($q = 1 \dots Q$) such that

$$x^l \bigcap x^m = \varnothing \tag{3}$$

and

$$\bigcup_{q=1}^{Q} x^q = x \tag{4}$$

Then let r_q the number of parameters in each group (clearly $\sum_{q=1}^{Q} r_q = n$) and each parameter group to be constrained by:

$$g(x) = \begin{cases} g^1(x_1, x_2, \ldots, x_{r_1}) = c^1 \\ g^2(x_{r_1+1}, x_{r_1+2}, \ldots, x_{r_1+r_2}) = c^2 \\ \ldots. \\ g^Q(x_{r_1+r_2+\cdots+r_{Q-1}+1}, \ldots, x_n) = c^Q \end{cases} \quad (5)$$

Let $x_{k_q} \in x^q$ denote the parameter which is chosen as dependent in each group. Then one has[5]*:*

$$df = \sum_{q=1}^{Q} \sum_{s=r_{q-1}}^{r_q} \left(f_s - f_{k_q} \frac{g_s^q}{g_{k_q}^q} \right) dx_s \quad (6)$$

where $f_s, g_s^q, f_{k_q}, g_{k_q}^q$ denote the partial derivatives of f, g w.r.t. x_s,x_{k_q} respectively.

Note that f_s, f_{k_q} are the usual (free) partial derivatives of $f(x)$ w.r.t. x_s, x_{k_q}. We write:

Definition 1.

$$f_{s|k_q} = f_s - f_{k_q} \frac{g_s^q}{g_{k_q}^q} \quad (7)$$

and call $f_{s|k_q}$ the constrained derivative of f w.r.t. x_s when x_{k_q} is the dependent parameter of constraint q.

For the purposes of this paper, it is particularly useful the following Corollary (the proof is in Appendix 1).

Corollary 1. *In the case of linear constraints, one gets:*

$$f_{s|k_q} = f_s - f_{k_q} \quad (8)$$

Theorem 1 and Corollary 1 define the rate of change of $f(x)$ in respect of x_i, in the presence of input constraints. In terms of our work, they answer Question 1.

The answer to Question 2, namely, how the variation of a model output is apportioned by the inputs, is found utilizing the following sensitivity measure (Borgonovo & Apostolakis, 2001; Borgonovo & Peccati, 2004):

$$D_s(x, dx) = \frac{df_s(x)}{df(x)} = \frac{f_s(x)dx_s}{\sum_{j=1}^{n} f_j(x)dx_j} \quad (9)$$

D is the ratio of the (infinitesimal) change in f caused by a change in x_s over the total change in f caused by a change in all the model parameters.

It can be shown that the following properties hold for D (Borgonovo & Apostolakis, 2001; Borgonovo & Peccati, 2004):

Property 1 Additivity. Let $\{x_i, x_j, \ldots, x_s\}$ and $S = (i, j, \ldots, k)$ the set of the corresponding indices. Then the joint sensitivity of f on contemporary changes in $\{x_i, x_j, \ldots, x_s\}$ is given by:

$$D_S(x,dx) = \frac{\sum_{s \in S} f_s(x) dx_s}{\sum_{j=1}^{n} f_j(x) dx_j} = \sum_{s \in S} D_s(x,dx) \tag{10}$$

Property 2 Sum to unity.

$$\sum_{s=1}^{n} D_s(x,dx) = 1 \tag{11}$$

Property 3 Let us consider uniform changes in the parameters, i.e.,

$$(H1) \quad dx_j = dx_k \ \forall j,k = 1,2,..,n \tag{12}$$

If (12) holds, then

$$D1_s = \frac{f_s}{\sum_{j=1}^{n} f_j} \tag{13}$$

which means that partial derivatives are proportional to D under an assumption of uniform changes in the parameters.

This implies that using partial derivatives to determine the influence of parameters on model output is equivalent to state the assumption that all parameters are varied by the same (small) quantity. This result has the practical consequence that, if two parameters have different dimensions, the corresponding partial derivatives cannot be compared (Borgonovo & Apostolakis, 2001; Borgonovo & Peccati, 2004, 2006). Thus, partial derivatives cannot be utilized to answer Question 2, in general.

Property 4 Let us consider proportional relative changes in the parameters, i.e.,

$$(H2) \quad \frac{dx_j}{x_j} = \frac{dx_k}{x_k} = \omega \ \forall j,k = 1,2,..,n \tag{14}$$

If (14) holds, then:

$$D2_s = \frac{f_s \frac{x_s}{f}}{\sum_{j=1}^{n} f_j \frac{x_j}{f}} = \frac{E_s}{\sum_{j=1}^{n} E_j} \tag{15}$$

where E_s is the well known definition of elasticity of f w.r.t. x_s. Equation (15) implies that ranking basic events with elasticity is equivalent to rank them with $D2$ (15), i.e., it is equivalent to stating the implicit assumption that all parameters are changed by the same proportion.

Remark 1. Properties 3 and 4 have the following geometric interpretation. Let us write $dx = tv$, where v is the unit vector indicating the direction of the change and t its magnitude. Equation (9) can be rewritten as:

$$D_s(x,v) = \frac{f_s(x) v_s}{\Delta f \cdot v} \tag{16}$$

In (16), the denominator is the directional derivative of f, an the numerator is the fraction of the directional derivative related to variable x_s. In other words, D does not only include the rate of change of f, but also the direction of change.

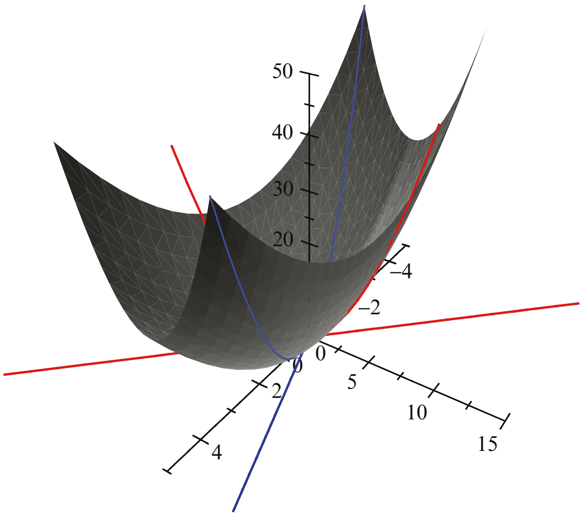


Fig. 2 2d Geometric interpretation of (13) and (15). The 45° line is represented by the *solid line*, the *dotted line* represents a line with tangent equal to ω ($\omega = -3$ in our plot)

To offer a visual interpretation, in two dimensions, (13) is equivalent to explore the behavior of a surface along a 45° line lying on the x–y plane, while (15) is equivalent to explore the sensitivity of a surface along a line with tangent equal to ω and lying on the x–y plane (Fig. 2).

We again note that the above definition of D does consider inputs as free. To be able to measure the importance of inputs in the case constraints apply, one needs to combine D, as defined by (9), and the constrained Sensitivity analysis results of Theorem 1. One gets:

Corollary 2. *The sensitivity measure for Question 2 in a constrained sensitivity is given by:*

$$D_{s|k_q}(x,v) = \frac{df_s}{df} = \frac{\left(f_s - f_{k_q}\frac{g_s^q}{g_{k_q}^q}\right)dx_s}{\sum_{q=1}^{Q}\sum_{l=r_{q-1}}^{r_q}\left(f_l - f_{k_q}\frac{g_l^q}{g_{k_q}^q}\right)dx_l} \tag{17}$$

The sensitivity measures for uniform and proportional changes follow straightforwardly and we do not report them.

The following remarks hold:

Remark 2. • Equations (7) and (17) imply that the sensitivity results depend on the parameter which is selected as the dependent one. They provide a differential explanation of the dependence effect registered in Example 1.

- Equation (8) implies that:

$$f_{k_q|k_q}(x) = 0 \tag{18}$$

what means that once a parameter is chosen as dependent, the sensitivity of any model output on this parameter is null.

- From (18) there follows that:

$$D_{k_q|k_q}(x,v) = 0 \ \ \forall v \tag{19}$$

Equation (19) implies that once a parameter is chosen as dependent one, the importance of the parameter is null, independently of the direction of change.

- Let x^q denote the set of parameters that must satisfy the qth constraint. From (10), one gets that the impact of x^q is given by:

$$D_{x^q|k_q}(x,v) = \frac{df_{s_q}}{df} = \frac{\sum_{s=r_{q-1}}^{r_q}\left(f_s - f_{k_q}\frac{g_s^q}{g_{k_q}^q}\right)dx_s}{\sum_{q=1}^{Q}\sum_{l=r_{q-1}}^{r_q}\left(f_l - f_{k_q}\frac{g_l^q}{g_{k_q}^q}\right)dx_l} \tag{20}$$

The next section defines the application of the above results to vNM models.

2.2 Sensitivity Analysis of vNM Models

The purpose of this section is to obtain analytical results for the sensitivity analysis (Questions 1 and 2) of vNM models. The section is divided into two parts. The first part lays down the notation and derives the analytical expression of a decision-support model output relying on vNM theory. The second part then derives analytical results by application of Theorem 1, and Corollary 2.

We try and follow the notation in Kreps (1988). In vNM models a decision-maker is selecting between A alternatives, to each alternative there can correspond Z state of natures, with given probability measure p_z, $\sum p_z = 1$, and utility u_z. We let $\mathbb{P}$ denote the set of all measures and p an element of $\mathbb{P}$. From vNM theorem, there follows directly that the utility of alternative i is an affine function of p_z:

$$u_i = f_i(p) = \sum_z p_z u_z \tag{21}$$

In applications (see Fig. 1 for an example), between a decision node and a state of nature (called consequence in decision analysis model implementation (see Celemen (1997)) there interpose several other events. We denote the generic chance node with q, the number of nodes with Q, and the number of outcomes of each node with r_q, $q = 1,2,\ldots,Q$ (Fig. 3). Then $n = \sum_{q=1}^{Q} r_q$ is the number of events in the model. We denote as E_s a generic event, which, in a tree or diagram, is a node outcome. Then, letting p_s, $s = 1,2,\ldots,n$, denote the corresponding conditional probability, at each node it holds that:

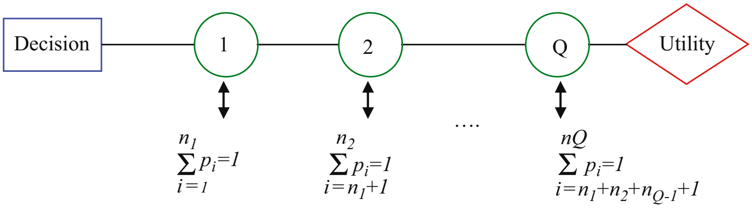


Fig. 3 Correspondence between the constrained sensitivity analysis problem and the VNM expected utility sensitivity analysis problem

$$\begin{cases} \sum_{s=1}^{r_1} p_s = 1 \\ \sum_{s=r_1+1}^{r_1+r_2} p_s = 1 \\ \cdots \\ \sum_{s=r_1+r_2+\cdots+r_{Q-1}+1}^{n} p_s = 1 \end{cases} \tag{22}$$

Let $E_{1_z^i}, E_{2_z^i}, \dots, E_{Q_z^i}$ denote the set of outcomes whose happening is necessary for z to be reached given that alternative i is selected. We call $E_{1_z^i}, E_{2_z^i}, \dots, E_{Q_z^i}$ a conditional path leading to consequence z, where the conditional here means given that alternative i is selected. Let $L_z^i = \{1_z^i, 2_z^i, \dots, Q_z^i\}$ the collection of the corresponding indices. Then, by construction,

$$p_z = P(E_{1_z^i}, E_{2_z^i}, \dots, E_{Q_z^i}) = \prod_{l \in L_z^i} p_l \tag{23}$$

One can then write the expected utility of alternative i, u_i, as the following affine function of p_l:

$$u_i = f_i(p) = \sum_z (\prod_{l \in L_z^i} p_l) u_z \tag{24}$$

According to vNM theory, the decision-maker selects alternative i that maximizes $f_i(p)$. Therefore the expected utility of a decision as a function of p is:

$$u = f(p) = \max_j f_j(p) = \max_j \sum_z (\prod_{l \in L_z^i} p_l) u_z \tag{25}$$

$f(p)$ is a piecewise-defined function of p as p varies in $\mathbb{P}$. In particular, $u = f_i(p)$ if alternative i is the preferred one at p. Now, let

$$\mathbb{P}_i = \left\{ p \in \mathbb{P} : i = \arg\max_i f_i(p) \right\} \tag{26}$$

denote the subset of p at which alternative i is preferred. Then, $\mathbb{P}_i \subseteq \mathbb{P}$ can be called preference region of alternative i (Borgonovo and Peccati, 2006). $\partial \mathbb{P}_i$ denotes the frontier of $\mathbb{P}_i$, and $\overline{\mathbb{P}_i}$ the closure of $\mathbb{P}_i$.

Now, as we stated in Sect. 1, we are interested in the properties of $u = f(p)$ as p varies in $\mathbb{P}$. The following holds for (25) (see Appendix 1 for the proof).

Proposition 1. *$u = f(p) \in \mathbb{C}^{\infty}(\mathbb{P}_i)$ but is at most $\mathbb{C}^{0}(\mathbb{P})$.*

In a vNM settings, the decision-maker knows exactly the point $p \in \mathbb{P}$ where He/She has to make the decision. For vNM problems, thus, the sensitivity analysis exercise is a local one, namely, one wants to appreciate the sensitivity of the expected utility at p. We write:

$$\begin{array}{c} Sensitivity\_of\_ \\ u = f(p) \\ with \\ \sum_{s=1}^{r_1} p_s = 1 \\ \sum_{s=r_1+1}^{r_1+r_2} p_s = 1 \\ \cdots \\ \sum_{s=r_1+r_2+\cdots+r_{Q-1}+1}^{n} p_s = 1 \end{array} \tag{27}$$

The sensitivity analysis problem of (27) is represented in Fig. 3, which illustrates the correspondence between event outcomes and sum to unity of the corresponding conditional measures.

Exploiting (25) and Theorem 1, we can then introduce the following result (the proof is in Appendix 1).

Theorem 2. *Let $p \in \mathbb{P}_i$, then:*

1. *The free partial derivative of the expected utility w.r.t. probability p_t ($t = 1,2,\ldots,n$) is given by:*

$$\frac{\partial}{\partial p_t} f(p) = \sum_{z \in Z_t} p_z^{t-} p_z^{t+} u_z \tag{28}$$

where (see Appendix 1)

$$\begin{array}{l} p_z^{t-} = \text{ is the probability of the events that precede } E_t \text{ in the path} \\ p_z^{t+} = \text{ is the probability of the events that follow } E_t \text{ in the path} \end{array} \tag{29}$$

2. *Let E_t be an outcome of node q. Let t_q be the pivotal probability. Then, the constrained partial derivatives of the expected utility w.r.t. probability p_t is:*

$$u_{t|t_q} = \frac{\partial}{\partial p_t} f(p) - \frac{\partial}{\partial p_{t_q}} f(p) = \sum_{z \in Z_t} p_z^{t-} p_z^{t+} u_z - \sum_{z \in Z_{t_q}} p_z^{t_q-} p_z^{t_q+} u_z \tag{30}$$

3. *The importance of a probability w.r.t. an expected utility in a vNM model is given by:*

$$D_{t|t_q}(p, dp) = \frac{\left(\sum_{z \in Z_t} p_z^{t-} p_z^{t+} u_z - \sum_{z \in Z_{t_q}} p_z^{t_q-} p_z^{t_q+} u_z \right) dp_s}{\sum_{j=1}^{n} \left(\sum_{z \in Z_j} p_z^{j-} p_z^{j+} u_z - \sum_{z \in Z_{j_q}} p_z^{j_q-} p_z^{j_q+} u_z \right) dp_j} \tag{31}$$

Let us now state a some observations.

Remark 3. From (30), it is immediate to find back the result that, in the case all events are binary, and E_{t-1} and E_t belong to the same node, then

$$\frac{\partial}{\partial p_t} f(p) = -\frac{\partial}{\partial p_{t-1}} f(p) \tag{32}$$

i.e., sensitivity of an expected utility on a probability is the opposite to its sensitivity to the complementary of that probability.

Remark 4. If one utilizes a normalized utility $u_z \in [0,1]$, then one obtains that $\frac{\partial}{\partial p_t} f(p) \geq 0$ (28). However, to draw the conclusion that an increase in any probability leads to an increase in expected utility would be an erroneous interpretation of the results. In fact, probabilities are not free to vary, but, as stated in (27), a change in one probability is reflected by a contemporary opposite change of all the other probabilities in the same constraint. Such an effect is contained in (30), which accounts for the bouncing back of the other probabilities through the term $\left(-\sum_{z \in Z_{t_q}} p_z^{t_q-} p_z^{t_q+} u_z\right)$.

Remark 5. If one were interested in the joint importance of a group of probabilities, then one would simply have to sum the importances of the probabilities in the group thanks to the additivity property.

Remark 6. Point 2 of Theorem 2 answers Question 1 for vNM models. Question 2 is answered by point 3.

We would like to observe that (30) can be utilized to derive indication on the model correctness. Through (30), in fact, an analyst is deriving information on the rate of change and relevance of the probabilities on one of the alternatives. The results can be utilized to infer, for example, information on the model correctness. If one is expecting an increase in utility and obtains the opposite effect, then one is lead to further investigate whether the discrepancy is due to false expectations or to some computational error.

Finally, in virtue of Proposition 1, $f(p)$ is not differentiable on the frontiers of alternative preference regions. Indeed, $f(p)$ is subdifferentiable. The one-sided derivatives at each $\overline{p} \in \partial\mathbb{P}_i$ are defined. This has the following decision-theoretic meaning. Suppose the decision-maker is indifferent among two alternatives, i and j (for simplicity), then $u_{t|t_q}(\overline{p})$ changes depending on whether the sensitivity is explored in a direction that from $\overline{p}$ leads to a point $p' \in \mathbb{P}_i$ or to a point $p'' \in \mathbb{P}_j$. But moving from $\partial\mathbb{P}_i$ to $\mathbb{P}_i$ or $\mathbb{P}_j$ has the meaning of making alternative i or j the preferred one. Thus, the answer to Questions 1 and 2 is unique provided that the decision maker selects one of the alternatives as preferred.

3 Sensitivity Analysis with Uncertainty: The Expected Value Problem

In many decision analysis applications once the model is developed and some base case values of the probabilities are assessed, the decision-maker starts questioning His/Her state of belief of p. Furthermore, in many decision-support models, probabilities are computed as functions of some uncertain parameters. Suppose that node q in the model of Fig. 3 contains two outcomes: E_t = "the equipment fails before T" and its complementary $\overline{E_t}$ (no failure). If, as it is usually the case in the practice (Borgonovo & Apostolakis, 2001) the failure time is modelled by an exponential distribution. Then:

$$P(E_t) = 1 - e^{-\theta T} = p_t(\theta) \tag{33}$$

If the decision-maker utilizes a Bayesian approach to estimate θ then, letting $h(\theta)$ denote the density of θ w.r.t. the Lebesgue measure, $P(E_t)$ has an induced density $f_{P(E_t)}(p_t) = \frac{1}{T(1-p_t)} h[-\frac{\ln(1-p_t)}{T}]$.

Before providing the sensitivity analysis results, we study the change in model structure, namely (24) and (25), implied by the subjective uncertainty framework. We adopt a standard Bayesian decision theoretic setting (Berger, 1985; Insua et al., 1998). Let $\theta \in \Theta \subset R^n$ be the state variable, or "*parameter*" (Gustafson and Wasserman, 1995), $\mathcal{B}$ a σ-algebra on Θ and μ the probability measure on $(\Theta, \mathcal{B})$ which reflects the current decision-maker's state of belief on the parameters. As a function of the parameters,[6] (24) becomes:

$$u_i = g_i(\theta) = \sum_z \Big(\prod_{l \in L_z^i} p_l(\theta)\Big) u_z \tag{34}$$

The expected utility of alternative i is, then, (Insua et al., 1998):

$$\mathbb{E}_\mu [g_i(\theta)] = \int_\Theta g_i(\theta) d\mu(\theta) \tag{35}$$

The decision-maker selects alternative i such that:

$$i = \arg \max_{j=1,2,\dots,A} \mathbb{E}_\mu [g_j(\theta)] \tag{36}$$

Equation (36) is standard in Bayesian statistics. We would like to offer an interpretation of such relation in a Anscombe and Aumann (1963) sense. Equation (36) can be read as follows. Let Θ play the role of Ω, the state of the world in Anscombe–Aumann's model. A state of the world is represented by a value $\theta \in \Theta$. Given θ, the decision-maker has to compare lotteries, $\mathcal{L}_j = \sum_z \Big(\prod_{t \in L_z^j} p_t(\theta)\Big) u_z$. The Anscombe-Aumann axioms, then, assure that

$$i \succeq j \Longleftrightarrow \mathbb{E}_\mu [g_i(\theta)] \geq \mathbb{E}_\mu [g_j(\theta)] \tag{37}$$

We are now ready to derive the following result (the proof is in Appendix 1).

Theorem 3. *Let* $p=p(\theta)$, *(*$\mathcal{B},\Theta,\mu$*) a probability space with* $d\mu(\theta)=\prod_{i=1}^{n} d\mu_i(\theta_i)$.

1.

$$u = f(\widehat{p}) = \max_j \sum_z \left(\prod_{l \in L_z^i} \widehat{p}_l \right) u_z \tag{38}$$

where

$$\widehat{p}_l = \mathbb{E}_\mu \left[p_l(\theta) \right] \tag{39}$$

2. The rate of change the expected utility w.r.t. due to a change in the expected value of probability p_l, $u_{t|t_q}$, *is given by:*

$$u_{t|t_q} = \sum_{z \in Z_t} \widehat{p}_z^{t-} \widehat{p}_z^{t+} u_z - \sum_{z \in Z_{t_q}} \widehat{p}_z^{t_q-} \widehat{p}_z^{t_q+} u_z \tag{40}$$

3. The contribution of a change in probability to the expected utility change is identified by:

$$D_{t|t_q}(\overline{p}, dp) = \frac{\left(\sum_{z \in Z_t} \widehat{p}_z^{t-} \widehat{p}_z^{t+} u_z - \sum_{z \in Z_{t_q}} \widehat{p}_z^{t_q-} \widehat{p}_z^{t_q+} u_z \right) dp_s}{\sum_{l=1}^{n} \left(\sum_{z \in Z_l} \widehat{p}_z^{l-} \widehat{p}_z^{l+} u_z - \sum_{z \in Z_{l_q}} \widehat{p}_z^{l_q-} \widehat{p}_z^{l_q+} u_z \right) dp_l} \tag{41}$$

From a Bayesian viewpoint the above theorem has the following interpretation. Suppose that some evidence happens, that causes a small deviation in the posterior expected value of the probabilities. Then the rate of change of the utility due to such evidence is given by (40) and the importance of the probability by (41). Note that such equations have the same form of (30) and (31), provided that one utilizes the expected value of the probabilities in (30) and (31). In other words, if the reference value of the probabilities in the vNM model coincides with their expected value in the Bayesian one, under the assumption of independent parameters, then the sensitivity analysis. Problems on small changes in the reference values of the probabilities lead to the same answers for a vNM decision-maker and a Bayesian one solving the expected value problem.

However, the above results hold only under the condition that the parameters are independent. To say it using Wallace (2000) terminology, in that case uncertainty is "*absorbed*" by the expected values of the parameters, and the deterministic method still maintains some value. However, if the independency assumption does not hold, the above results would reveal to be unsatisfactory. We discuss this problem in the next section.

4 Sensitivity Analysis with Uncertainty: The State of Belief Problem

We are now left to discuss the sensitivity analysis problem of finding the parameter that influence the decision analysis model response the most when μ is not a product measure (no independence as opposite of the previous section).

If is clear that, to answer the above question, the local approach discussed for vNM models is not appropriate anymore, neither is Theorem 3. A traditional sensitivity measure to quantify the relevance of a parameter, say θ_s, is the expected value of sample information, defined as (see Pratt et al. (1995)):

$$EVSI_s = \mathbb{E}_{\mu(\theta|\theta_s)}[\max_j g_j(\theta)] - \max_{j=1,2,\ldots,A} \mathbb{E}_\mu [g_j(\theta)] - c \tag{42}$$

where $\mu(\theta|\theta_s)$ is the conditional measure given θ_s and c is the cost of testing. As usual, we note the exchange between the max and expectation operations in (35). Hence, if one takes as importance measure of θ_s its $EVSI$, one is ranking uncertain parameters based on the expected gain in utility that follows from obtaining information on the parameter. Borgonovo and Peccati (2006) discuss that under the assumption of a square loss function, the EVSI of a parameter is proportional to $\mathbb{V}^s_{\mu(\theta)}[\max_j g_j(\theta|\theta_s)]$, the contribution of the parameter to $\mathbb{V}_{\mu(\theta)}[\max_j g_j(\theta|\theta_s)]$, and the decomposition of $\mathbb{V}_{\mu(\theta)}[\max_j g_j(\theta|\theta_s)]$ is obtained by means of Efron and Stein (1981) result, together with the methodology of Sobol' (1993).

Alternatively, one may want to measure the influence of a parameter without relying on any of the moments of the utility function (the possibility of normalizing and exchanging utility and probability is thoroughly discussed in Castagnoli and Li Calzi (1996)). A sensitivity measure that does not look at any moment of $g_j(\theta)$, but considers the entire (distribution of) the utility can be built as follows. Consider $g_j(\theta)$. As θ has measure μ, then $g_j(\theta)$ as a function of random variable inherits an induced distribution μ_{U_j}. If θ_s does not affect the decision-maker view of the problem at all, then the utility is independent of θ_s, and its distribution is not affected by θ_s assuming any of its values. In other words, $\mu_{U_j} = \mu_{U_j|\theta_s}$. However, if θ_s influences $g_j(\theta)$, then $\mu_{U_j} \neq \mu_{U_j|\theta_s}$. Then, let us define the following quantity:

Definition 2.

$$\delta_s = \frac{1}{2}\mathbb{E}_{\theta_s}\left[\int \left| d\mu_{U_j} - d\mu_{U_j|\theta_s} \right|\right] \tag{43}$$

The interpretation of δ_s becomes clearer if one allows for density $f_{Uj}(u)$ associated with μ_{U_j} and the conditional densities $f_{U_j|\theta_s}(u)$ associated with $\mu_{U|\theta_s}$. Then one can write:

$$\delta_s = \frac{1}{2}\mathbb{E}_{\theta_s}\left[\int \left| f_{U_j}(u) - f_{U_j|\theta_s}(u) du \right|\right] \tag{44}$$

The quantity inside the expectation operator, namely,

$$\varsigma(\theta_s) = \int \left| f_{U_j}(u) - f_{U_j|\theta_s}(u) \right| du \tag{45}$$

Table 1 Properties of the delta uncertainty importance measure

No.	Property
1	$0 \leq \delta_s \leq 1$
2	$\delta_s = 0$ if $g_j(\theta_s)$ is independent of θ_s
3	$\delta_{1,2,\ldots,n} = 1$
4	$\delta_{sl} = \delta_s$ if $g_j(\theta)$ is dependent on θ_s but independent of θ_l
5	$\delta_s \leq \delta_{sl} \leq \delta_s + \delta_{l\|s}$

measures the shift which is provoked in the utility when θ_s is fixed at one of its possible values. Geometrically, $\varsigma(\theta_s)$ is the area between the densities $f_{U_j}(u)$ and $f_{U_j|\theta_s}(u)$ (Borgonovo, 2006a, 2007). $\varsigma(\theta_s)$ is dependent on θ_s, and, as such, $\varsigma(\theta_s)$ is a function of random variable. Taking the expectation based on the marginal distribution of θ_s, namely $\mathbb{E}_{\theta_s}[\varsigma(\theta_s)]$ (43), one measures the average shift in the decision-maker utility provoked by θ_s.

The definition of δ can be extended to any group of parameters, $\underline{R} = (\theta_{i_1}, \theta_{i_2}, \ldots, \theta_{i_r})$, as follows:

$$\delta_{i_1,i_2,\ldots,i_r} = \frac{1}{2}\mathbb{E}_{\underline{\mu_R}}[\varsigma(\underline{R})] \tag{46}$$

Given the definitions in (43) and (46), δ shares the properties reported in Table 1 (the proofs can be found in Borgonovo (2006a)). One can summarize these properties as follows. Property 1 implies that the δ of an individual parameter or of a group can only assume values between 0 and 1. Property 2 suggests that a parameter/group has null importance when $g_j(\theta)$ is independent of that parameter/group. Property 3 states that the joint importance of all inputs equals 1. Properties 4 and 5 refer to the joint importance of two (or more) parameters. Property 4 says that if $g_j(\theta)$ is dependent on θ_s but independent of θ_l then $\delta_{sl} = \delta_s$. Property 5 states that the joint importance of two parameters is greater than the importance of an individual parameter, but limited by the sum of such importance and the conditional term $\delta_{l|s}$ given by (see also Borgonovo (2006a, 2007)):

$$\delta_{l|s} = \frac{1}{2}\mathbb{E}_{\theta_s\theta_l}\left[\int \left|f_{U|\theta_s}(u) - f_{U|\theta_s\theta_s}(u)\right| du\right] \tag{47}$$

We would like to state some observations.

Remark 7. The definitions of the δ uncertainty importance for individual parameters (43) and for groups (46) do not require parameter independence.

Remark 8. A decision-maker utilizing EVSI or δ in a sensitivity analysis, is gaining information on Question 2, as She/He is capable of assessing the relative influence of the parameters. Note that, since EVSI and δ are defined in respect of the current state of belief, both EVSI or δ do not provide information on the direction of change in the expected utility given a change in the decision-maker state of belief (Question 2).

The answer to Question 1 for a Bayesian decision-maker requires a different approach. Namely, one needs to solve the problem:

Problem 1. What is the effect on $\mathbb{E}_{\mu}[g_i(\theta)]$ given a change in the decision-maker's state of belief?

Note that Problem 1 is the reformulation of Question 1 in a Bayesian setting. The answer to Problem 1 finds its statistical foundations lie in Bayesian statistics for the sensitivity of posterior expectations, and in particular on the theoretical framework developed in Gustafson (1996). Let $\pi(\theta)$ the density associated with μ. The change in the decision-maker's state of belief is represented by a change in $\pi(\theta)$. The change in $\pi(\theta)$ is represented by a perturbation $\varepsilon(\theta)$ that leads to the new state of belief density $w(\theta) = \pi(\theta) + \varepsilon(\theta)$. The technical meaning and properties of a perturbation are discussed in Gustafson (1996). The fundamental idea is that the density $\pi(\theta)$ is distorted through the new term $\varepsilon(\theta)$. Now, consider the expected value functional $\mathbb{E}_{\pi}[g_i(\theta)]$, where, we have evidenced π, the density associated with μ and representing the decision-maker's current state of belief. Then, it is possible to see that thanks to the linearity of $\mathbb{E}_{\pi}[g_i(\theta)]$ (Appendix 1),

Proposition 2. *The rate of change of the expected utility in the direction of change u is:*

$$D_{\pi}^{g_i}u = \int_{\Theta} g_i(\theta)\varepsilon(\theta)d\mu(\theta) \tag{48}$$

Equation (48) expresses the sensitivity of the expected utility in alternative i, given a change in the decision-maker's view on θ. With this respect, it is an answer to Question 1 and it parallels the use of (30) in the case of vNM models. However, some remarks have to be made.

The first concerns the effect of independence. In the case the decision-maker view of the problem is such that independence among the θ_s holds, then one can assess perturbations for each of the parameters separately. Thus, through $D_{\pi}^{g_i}u(\theta_s)$ (48) one is indeed answering the question of what is the effect of a change in the decision-maker's view on each θ_s. In the case correlations are present, then it is not possible to perturb each parameter independently, but one unique perturbation reflecting the decision-maker's change in view must be adopted. This has the following consequence. If the parameters of interests are probabilities, then the presence of the constraints (22), imply that the p's cannot be independent of each other and therefore assessing the sensitivity of the expected utility on each p_s separately is an ill-defined problem. In fact, what is concerned by the change is not the view on p_s, but in the subjective conditional measure of all the events associated with p_s.[7] This is the Bayesian equivalent of the constrained results obtained for the vNM decision-makers in Theorem 2.

The second concerns the utilization of (48). Equation (48) provides the decision-maker with the response of the expected utility of alternative i to a change in its state of belief. Thus, as mentioned, it addresses a different problem than the one of providing information on whether, in correspondence of the state of belief change, a change in the preferred alternative happens. Such a problem finds its answer in the works of, among others, Evans (1984), Fishburn et al. (1968), and Ringuest (1997), but also in Insua et al. (1998).

5 Conclusions

Ringuest (1997) maintains that "*any effective decision analysis must include a thorough sensitivity analysis*". Although sensitivity analysis is often regarded as an integral step of the decision-making process, many questions remain open concerning the consistency of the sensitivity analysis method in respect of the decision problem at hand.

We have formulated our analysis as an answer to two questions. The first is Samuelson's statement that has originated comparative statics. The question concerns the response of a system to (small) changes in the inputs; the answer is found through implicit differentiation and delivers the rate of change in the output given a small change in the inputs. The second question relates to the problem of how the change is apportioned to input variations (Tarantola, 2000; Saltelli, 2002). The answer to the second question defines the importance of parameters (Saltelli, 2002).

We have discussed the formulation of Questions 1 and 2 in decision-making. We have seen that, to answer these questions in a decision-theory consistent way, two completely different mathematical approaches must be undertaken depending on whether a decision-maker is utilizing a support model based on the vNM axioms or a model consistent with a Bayesian subjective utility approach. Table 2 offers a comparison of the results of this work.

We illustrate first the results that we have obtained for vNM models. We have seen that, since the analysis takes place at p – the given value of the probabilities – the answer to the two sensitivity analysis questions can be found by using differential techniques: partial derivatives for Question 1 – along the line of the comparative statics tradition – and the differential importance measure (D) for Question 2. Both measures relate to the infinitesimal change dp. However, the presence of the normalization condition turns the analysis into a constrained sensitivity problem. We have then illustrated that the solution of the problem is found by applying a recent result introducing the concept of constrained derivative. This has allowed us to come

Table 2 Comparison of sensitivity analysis results for a VNM and a Bayesian decision-maker

Type	Answer to Question 1	Answer to Question 2
vNM	$u_{t\|t_q} = \sum_{z\in Z_t} p_z^{t-} p_z^{t+} u_z - \sum_{z\in Z_{t_q}} p_z^{t_q-} p_z^{t_q+} u_z$	$D_{t\|t_q} = \frac{\left(\sum_{z\in Z_t} p_z^{t-} p_z^{t+} u_z - \sum_{z\in Z_{t_q}} p_z^{t_q-} p_z^{t_q+} u_z\right) dp_s}{\Sigma_{j=1}^n \left(\sum_{z\in Z_j} p_z^{j-} p_z^{j+} u_z - \sum_{z\in Z_{j_q}} p_z^{j_q-} p_z^{j_q+} u_z\right) dp_j}$
Bayesian	$D_\pi^{g_i} u = \int_\Theta g_i(\theta)\varepsilon(\theta) d\mu(\theta)$	$EVSI_s$ (42), δ_s (43)

to the the analytical expression of the sensitivity measures of vNM expected utility models both for Question 1 (comparative statics) and Question 2 (differential importance). We have then discuss the meaning and insights a decision-maker derives from application of the results.

In the practice of modeling, it is customary that, if time allows it, after preforming a sensitivity analysis at the base case, decision-makers start assessing uncertainty in the probabilities. In so doing, they leave the vNM framework and enter the Bayesian one. We have then analyzed how to find the answers for Questions 1 and 2 when the decision-maker is utilizing a model consistent with a (Bayesian) subjective expected utility approach. We have cast the sensitivity in a Bayesian framework and introduced the parameter space, Θ. We have seen that the answer to Question 1 is found by making use of a Bayesian sensitivity approach based on Fréchet differentiation (Table 2). More in detail, the change needs to be modeled as a perturbation of the decision-maker's density. We have derived the rate of change of the expected utility of an alternative as a function of the change (perturbation) in the decision-maker state of belief. We have seen that Question 2, namely the importance of parameter θ_s, is answered by means of either the $EVSI_s$ or by δ_s, if the decision-makers wants to assess the sensitivity of the entire distribution of the utility without reference to one of its moments (Table 2).

As a result of the work, one can conclude that, in general, a vNM decision-maker and a Bayesian one need to use very different approaches to answer the same sensitivity analysis questions. We have also looked for a bridge between the two approaches, called "the expected value problem". Namely, we have seen that sensitivity analysis results of a vNM decision-maker are the same as the ones utilized by a Bayesian decision-maker, if uncertainty is absorbed in the expected value. More precisely, if the Bayesian decision-maker's state of belief is such that the parameters are independent, and if the expected value of the probabilities is utilized by a vNM decision-maker as reference value, then, if the Bayesian decision-maker measures sensitivity as rate of change of the expected utility w.r.t. the expected value of the probabilities, the two decision-makers would derive the same results.

We note that the presented framework for sensitivity analysis paves the way to future research. First of all the answer to Questions 1 and 2 when these are posed in the context of models based on ambiguity theory and dynamic choice models. Further, among the above mentioned methods, the last approach, namely the perturbation approach, has not been thoroughly studied in applications. Thus, the exploration of issues such as the selection of the appropriate perturbation (partly discussed also in Gustafson (1996)) in the presence and absence of input correlations, is part of future work of the authors.

Notes

[1]We wish to point out the difference between the stability problem and the ambiguity problem. In the works relating to stability, one is interested in the sensitivity of a subjective expected utility to changes in the probability distributions. In other words, one is interested in the question of

whether selecting, say, a different prior the decision problem solution would change significantly, but one still remains in the realm of subjective expected utility. In the theory of decision making with ambiguity, originated by the Ellsberg paradox (Ellsberg, 1967),"*models have been proposed which extend subjective expected utility* (Ghirardato, Maccheroni, & Marinacci, 2004)" theory.

[2]Indeed, one can observe that fixing the inputs at p_0 is the same as having the decision maker assigning a delta-Dirac measure centered at p_0.

[3]If one tries to perform these operations on a standard software (for example DATAPRO), one would get an error message.

[4]See the previous footnote.

[5]By convention $r_0 = 1$.

[6]Nothing forbids the probabilities themselves to be directly the parameters.

[7]They are the events whose probabilities lie in the same constraint as p_s and are usually outcomes of the same node in the model.

Appendix 1: Proofs

Proof. of Corollary 1. Let the constraint be written as

$$g(x) = \left\{ \begin{array}{c} \sum_{s=1}^{r_1} x_s = c^1 \\ \sum_{s=r_1}^{r_1+r_2} x_s = c^2 \\ .... \\ \sum_{s=r_1+r_2+\cdots.+r_{Q-1}+1}^{n} x_s = c^Q \end{array} \right| \tag{49}$$

then

$$g_s^q(x) = g_{k_q}^q(x) = 1 \; \forall s, k_q = 1, 2, \ldots, n, \forall q = 1, 2, \ldots, Q \tag{50}$$

and the thesis follows from substitution in (7). □

Proof. of Remark 1. First let $\partial\mathbb{P}_i$ denote the frontier of an indifference region. If $p^0 \in \mathbb{P}_i/\partial\mathbb{P}_i$ (i.e., p^0 is an internal point), it holds that:

$$u = f_i(p^0) \tag{51}$$

Now, from (24) and (25), $f_i(p^0)$ is linearly multiplicative in p^0 and therefore $u \in \mathbb{C}^\infty(p^0)$. If $p^0 \in \partial\mathbb{P}_i$, then the decision-maker is indifferent between alternatives i and $j, \ldots, k$, and, by definition of indifference $f_i(p^0) = f_j(p^0) = \ldots = f_k(p^0)$, i.e., $f(p^0)$ is continuous when crossing indifference hypersurfaces. However, it easy to verify (it suffices an example) that u is not differentiable at $\partial\mathbb{P}_i$. □

Proof. of Theorem 2. Point 1.

1. If $p \in \mathbb{P}_i$, then from (25) and (24), we have:

$$u = u_i(p) = \sum_z \left(\prod_{l \in L_z^i} p_l \right) u_z = f(p) \tag{52}$$

The next step is to isolate the paths containing the probability of interest, p_t. To do so, it is necessary to group the paths containing event E_t. Let us denote this group as Z_t. Then, (52) can be rewritten as:

$$f(p) = \sum_{z \in Z_t} \left(\prod_{l \in L_z^i} p_l \right) u_z + \sum_{z \notin Z_t} \left(\prod_{l \in L_z^i} p_l \right) u_z \tag{53}$$

The terms $\sum_{z \notin Z_t} \left(\prod_{l \in L_z^i} p_l \right) u_z$ do not contain p_t and therefore they drop out in the differentiation. One can then concentrate on $\sum_{z \in Z_t} \left(\prod_{l \in L_z^i} p_l \right) u_z$. Noting that each summand is of the form:

$$\prod_{l \in L_z^i} p_l u_z = \left(\prod_{\substack{l<t \\ l \in L_z^i}} p_l \right) p_t \left(\prod_{\substack{l>t \\ l \in L_z^i}} p_l \right) \tag{54}$$

i.e., one can partition $\prod_{l \in L_z^i} p_l$ into three parts: 1 - $\left(\prod_{\substack{l<t \\ l \in L_z^i}} p_l \right)$ is the product of the probabilities of outcomes that precede E_t, 2 - p_t itself and 3 - $\left(\prod_{\substack{l>t \\ l \in L_z^i}} p_l \right)$ which is probability of events following E_t in the path. Therefore:

$$\frac{\partial u}{\partial p_t} = \sum_{z \in Z_t} \left(\prod_{\substack{l<t \\ l \in L_z^i}} p_l \right) \left(\prod_{\substack{l>t \\ l \in L_z^i}} p_l \right) u_z \tag{55}$$

Now, letting $p_z^{t-} = \left(\prod_{\substack{l<t \\ l \in L_z^i}} p_l \right)$ and $p_z^{t+} = \left(\prod_{\substack{l>t \\ l \in L_z^i}} p_l \right)$, one gets:

$$\frac{\partial u}{\partial p_t} = \sum_{z \in Z_t} p_z^{t-} p_z^{t+} u_z \tag{56}$$

Furthermore, by construction, p_z^{t-} is the probability of all outcomes leading to E_t, and p_z^{t+} the conditional probability of all outcomes following E_t and leading to consequence z.

2. One needs to combine Point 1 with Theorem 1
3. Combine (30) with (17). □

Proof. of Theorem 3. Given (34), the linearity of the expectation operator leads to:

$$\mathbb{E}_\mu\left[g_i(\theta)\right] = \sum_z \mathbb{E}_\mu\left[\prod_{l\in L_z^i} p_l(\theta)\right] u_z \tag{57}$$

and the assumption $d\mu(\theta) = \prod_{i=1}^n d\mu_i(\theta_i)$ leads to:

$$\mathbb{E}_\mu\left[\prod_{l\in L_z^i} p_l(\theta)\right] = \prod_{l\in L_z^i} \mathbb{E}_\mu\left[p_l(\theta)\right] \tag{58}$$

Thus:

$$\mathbb{E}_\mu\left[g_i(\theta)\right] = \sum_z \left(\prod_{l\in L_z^i} \widehat{p}_l\right) u_z \tag{59}$$

which leads to

$$u = \max_j \sum_z \left(\prod_{l\in L_z} \widehat{p}_l\right) u_z \tag{60}$$

where

$$\widehat{p}_l = \mathbb{E}_\mu\left[p_l(\theta)\right] \tag{61}$$

Points 2 and 3 follow from the fact that (38) is of the same form as (24) and follow from application of Theorem 2. □

Proof. of Proposition 48. Applying the definition of Fréchet differential to perturbations as in Gustafson (1996), then one has:

$$D_\pi^g u = \lim_{h\downarrow 0} \frac{\int g(\theta)\left[\pi(\theta) + h\varepsilon(\theta)\right] d\mu(\theta) - \int g(\theta)\left[\pi(\theta)\right] d\mu(\theta)}{h} = \int g(\theta)\varepsilon(\theta) d\mu(\theta)$$

□

Acknowledgments The authors wish to thank John Hey, Mohammed Abdellaoui, and the participants of the FUR 2006 conference for insightful comments on an earlier version of this work.

References

Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annual of Mathematical Statistics, 34*, 199–205

Berger, J. O. (1984). The robust Bayesian viewpoint (with discussion). In J. Kadane (Ed.), *Robustness in Bayesian statistics*. Amsterdam: Nother Holland

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Berlin: Springer

Berger, J. O. (1990). Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference, 25*, 303–328

Borgonovo, E. (2006a). Measuring uncertainty importance: Investigation and comparison of alternative approaches. *Risk Analysis, 26*(5), 1349–1361

Borgonovo, E. (2006b). *Comparative statics for multinomial trees* (internal memorandum). Milan, Italy: IMQ, Bocconi University

Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering and System Safety, 92*, 771–784

Borgonovo, E. (2008). Differential importance and comparative statics: An application to inventory management. *International Journal of Production Economics, 111*(1), 170–179

Borgonovo, E., & Apostolakis, G. E. (2001). A new importance measure for risk-informed decision-making. *Reliability Engineering and System Safety, 72*(2), 193–212

Borgonovo, E., & Peccati, L. (2004). Sensitivity analysis in investment project evaluation. *International Journal of Production Economics, 90*, 17–25

Borgonovo, E., & Peccati, L. (2006). On the quantification and decomposition of uncertainty. In: Uncertainty and Risk, M. Abdellaoui, R. D. Luce, M. Machina, & B. Munier (Eds.), *Theory and Decision Library Series*, Springer (Berlin), pp. 41–59, ISBN 978-3-540-48934-4

Bushena, D. E., & Zilberman, D. (1999). Testing the effects of similarity on risky choice: Implications for violations of expected utility. *Theory and Decision, 46*, 251–276

Castagnoli, E., & Li Calzi, M. (1996). Expected utility without utility. *Theory and Decision, 41*, 281–301

Celemen, R. T. (1997). *Making hard decisions: An introduction to decision analysis* (2nd ed.). Pacific Grove, CA: Duxbury Press, ISBN: 0534260349

Efron, B., & Stein, C. (1981). The Jackknife estimate of variance. *The Annals of Statistics, 9*(3), 586–596

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

Eschenbach, T. G. (1992). Scatterplots versus Tornado diagrams for sensitivity. *Interfaces, 22*(6), 40–46

Evans, J. R. (1984). Sensitivity analysis in decision theory. *Decision Sciences, 15*, 239–247

Fishburn, P. C., Murhpy, A. H., & Isaacs, H. H. (1968). Sensitivity of decisions to probability estimation errors: A reexamination. *Operations Research, 16*, 254–267

Ghirardato, P., Maccheroni, F., & Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory, 118*, 133–173

Gustafson, P. (1996). Local sensitivity of posterior expectations. *The Annals of Statistics, 24*(1), 174–195

Gustafson, P., & Wasserman, L. (1995). Local sensitivity diagnostics for Bayesian inference. *The Annals of Statistics, 23*(6), 2153–2167

Howard, R. A. (1988). Decision analysis: Practice and promise. *Management Science, 34*, 679–695

Insua, D. R., Ruggeri, F., & Martin, J. (1998). Bayesian sensitivity analysis: A review. *Proceedings of the Second International Symposium on Sensitivity Analysis of Model Output*, Venice, Italy, 1998, 255–257

Kreps, D. (1988). Notes on the Theory of Choice, *Underground Classics in Economics*, Westview Press Inc., Colorado (USA), ISBN: 08133755533

Pauwels, N., Van De Walle, B., Hardeman, F., & Soudan, K. (2000). The implications of irreversibility in emergency response decisions. *Theory and Decision, 49*, 25–51

Pratt, J. W., Raiffa, H., & Schlaifer, R. (1995). Introduction to statistical decision theory (904 pages). Cambridge, MA: The MIT Press, ISBN: 0262161443

Ringuest, J. L. (1997). L_p-metric sensitivity analysis for single and multi-attribute decision-analysis, *European Journal of Operational Research*, *98*, 563–570

Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Analysis, 22*(3), 579

Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modelling. *Statistical Science, 19*(4), 377–395

Samuelson, P. (1947). *Foundations of economic analysis*. Cambridge, MA: Harvard University Press

Smith, J. E. (1998). Evaluating income streams: A decision analysis approach. *Management Science, 44*(12), 1690–1708

Smith, J. E., & von Winterfeld, D. (2004) Decision analysis in management science. *Management Science, 50*(5), 561–574

Sobol, I. M. (1993). *Sensitivity estimates for nonlinear mathematical models*, Matem. Modelirovanie, 2(1) (1990) 112–118 (in Russian). English Transl.: Mathematical Models and Computational Experiments, 1(4) (1993), pp. 407–414

Tarantola, S. (2000). Quantifying uncertainty importance when inputs are correlated. In M. P. Cottam, R. P. Pape, D. W. Harvey, & J. Tait (Eds.), *Foresight and precaution*. Rotterdam: Balkema Editions, ISBN: 9058091406

Van Groenendaal, W. J. H. (1998). Estimating NPV variability for deterministic models. *European Journal of Operational Research, 107*, 202–213

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed., 648 pages). Princeton, NJ: Princeton University Press, reprinted May 1, 1980, ISBN-10: 069100362

Wallace, S. W. (2000). Decision making under uncertainty: Is sensitivity analysis of any use? *Operations Research*, (1), 20–25

Alternation Bias and the Parameterization of Cumulative Prospect Theory

K. Kaivanto

Keywords: Cumulative prospect theory · St. Petersburg Paradox · Local representativeness effect · Alternation bias · Law of small numbers

1 Introduction

Several authors have recently addressed the question of whether cumulative prospect theory (CPT) resolves the St. Petersburg Paradox (Blavatskyy, 2005; Rieger & Wang, 2006). These authors show that *direct application* of CPT to the St. Petersburg gamble fails to resolve the paradox under most conventional CPT parameterizations. They also propose a number of remedial fixes to CPT, central among which is a constraint on the value function exponent to be smaller than the probability weighting function exponent $(\alpha < \gamma)$. As this constraint is violated by most experimentally determined CPT parameterizations,[1] the remedy amounts to a fundamental reparameterization of CPT.

Tversky and Kahneman's (1992) CPT is a descriptive theory. It is consistent with stochastic dominance and accounts for framing effects, nonlinear probability preferences,[2] source dependence, risk seeking behavior,[3] loss aversion,[4] and uncertainty aversion. Nowhere has it been suggested that CPT's descriptive power extends to local representativeness effects. As Tversky and Kahneman (1992) stress, "Theories of choice are at best approximate and incomplete." Like numerous other heuristics, the operation of the representativeness heuristic (Tversky & Kahneman, 1974) depends in a complex fashion not only on the structural formulation of the decision problem, but also on its context and manner of presentation. Hence, explicit incorporation of the representativeness heuristic into the formal structure of CPT

K. Kaivanto
Department of Economics, Lancaster University Management School, Lancaster LA1 4YX, UK
e-mail: k.kaivanto@lancaster.ac.uk

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.

© Springer-Verlag Berlin Heidelberg 2008

would limit its applicability to a narrow range of problems. Incorporation of the binary sequence variant of the representativeness heuristic, the local representativeness heuristic, into the formal structure of CPT would limit its applicability still further. Thus, in order for CPT to function as a straightforwardly implementable general purpose descriptive theory, the local representativeness effect must remain outside its formal structure and specification.

Nevertheless, it is evident that the coin-tossing sequence found in the St. Petersburg gamble is precisely the sort of context where the local representativeness effect is likely to be operative. Indeed many psychological studies of randomness perception and the local representativeness effect in particular have utilized coin-tossing experiments (e.g. Rapoport & Budescu, 1997; Kareev, 1995). Moreover, these experimental studies of local representativeness have been designed in such a fashion[5] so as to allow 'clean' estimates of the alternation bias – i.e. estimates that are free from the confounding of conditional probability distortion with outcome value weighting.

Under the alternation bias, subjects perceive negatively autocorrelated sequences as maximally random, while the runs that are characteristic of unbiased memoryless Bernoulli processes are perceived as being excessively regular to be random. Therefore alternation bias leads to the subjective association of a *negative autocorrelation* with known *memoryless* and unbiased Bernoulli processes. This may be viewed as a subjective distortion of conditional probability. As phenomena ranging from the Gambler's Fallacy[6] to behavior in the Monty Hall problem[7] attest, people without specialist training in probability theory generally process conditional probability information differently than probability calculus intimates.

This note contends that once alternation bias is controlled for, conventional parameterizations of CPT do indeed succeed in resolving the St. Petersburg Paradox. The suggestion, made by Blavatskyy (2005) and Rieger and Wang (2006), to constrain the value function exponent to be smaller than the probability weighting function exponent $(\alpha < \gamma)$, confounds the subjective distortion of conditional probability with (a) the subjective distortion of unconditional probability and (b) the subjective valuation of outcomes. Reparameterization of CPT on the basis of the St. Petersburg Paradox is not only unnecessary, but would also disturb the theory's internal consistency and narrow its scope of applicability.

This note builds upon insights derived from Rabin (2002) on local representativeness. The following section briefly summarizes the local representativeness effect and presents empirical estimates of fist-order and higher-order alternation bias. Section 3 uses these estimates to show: (Sect. 3.1) that alternation bias on its own is sufficient for the subjective (mathematical) expectation of the St. Petersburg gamble to be rendered finite and within conventionally accepted empirical bounds; (Sect. 3.2) that alternation bias relaxes the Blavatskyy–Rieger–Wang CPT finiteness constraint; and (Sect. 3.3) that once alternation bias is controlled for, CPT yields a finite willingness to pay for the St. Petersburg gamble, which moreover falls within conventionally accepted empirical grounds. Section 4 concludes.

2 Local Representativeness Effect

That people display alternation bias in sequential randomization tasks was first hypothesized by Reichenbach (1934).[8] Experimental and observational evidence consistent with this hypothesis amassed from the time that the hypothesis was first put to test. The effect is known by different labels in different contexts – such as the Gambler's Fallacy in gambling, and the alternation bias in coin tossing – but these are specific examples of the local representativeness effect.[9]

People under the influence of the local representativeness effect attribute the salient properties of the population or generating process to short sequences. That is, such individuals do not adequately distinguish between local features and the properties of the whole, and they apply the latter to the former. For Bernoulli sequences this translates into local matching of outcome proportions with those of the long-run process, i.e. $(0.5, 0.5)$ for unbiased coins, and excessive local irregularity, i.e. a propensity to anticipate too many 'reversals' in short series.

This subjective predisposition to anticipate reversals is called the alternation bias, which equates to a negative subjective autocorrelation whereby people expect too few streaks in random sequences. Alternation bias effects have been documented up to sixth order (Budescu, 1987). Most empirical studies place the first-order effect at $P(H|T) = 0.6$ (see Bar-Hillel and Wagenaar (1991), Budescu (1987), Kareev (1995)). Still, some studies have found an even stronger first-order effect of $P(H|T) \in [0.7, 0.8]$ (Gilovich, Vallone, & Tversky, 1985). Taking higher-order effects into consideration, Rabin (2002) derives the following conditional probabilities from data presented in Rapoport and Budescu (1997): $P(H|T) = 0.585$, $P(H|HT) = 0.46$, $P(H|HHT) = 0.38$ and $P(H|HHH) = 0.298$, where this last expression refers to the conditional probability of a toss turning up 'Heads' given that the three immediately preceding tosses turned up 'Heads'. After rounding we obtain the following higher-order transition probabilities (see Table 1) with an alternation bias that bridges, between first and third orders, the weaker and the stronger alternation bias magnitudes reported in the literature.

3 Application to the St. Petersburg Gamble

In the modern variant of the St. Petersburg Paradox, a subject is offered a stochastic payout of $2^{\tilde{n}}$ dollars, where $\tilde{n}$ is the index of the first toss on which a fair, memoryless coin turns up 'Heads'.[10] The paradox arises because although

Table 1 Alternation bias estimates from the literature expressed as transition probabilities

First order	Higher order		
$P(H\|T)$	$P(H\|T)$	$P(H\|TT)$	$P(H\|TTT)$
0.6	0.58	0.62	0.70

the mathematical expectation of the gross St. Petersburg gamble payout $G_{StP} = 2^{\tilde{n}}$ is infinite, people are typically willing to pay only a small, finite amount to obtain this gamble. The theoretical literature favors Willingness To Pay (WTP) estimates between \$2 and \$4. This accords with Bernoulli's (1738) 'expected moral worth' solution which he formalized using the logarithmic function: abstracting from prior wealth, the St. Petersburg gamble is evaluated as $E[u(G_{StP})] = \sum_{n=1}^{\infty} 2^{-n} \log(2^n) = \sum_{n=1}^{\infty} \frac{n}{2^n} \log(2) = 2\log(2) = \log(4)$, giving a certainty equivalent of \$4 (Schmeidler & Wakker, 1998).[11] With the exception of Bottom, Bontempo, and Holtgrave (1989), formal experimental studies of the St. Petersburg gamble are thin on the ground, if for no other reason than the difficulty of bankrolling potentially very large (infinite in expectation) payouts. Some sources report that the typical WTP is no more than \$10 (Chernoff & Moses, 1959), while others, such as Camerer (2005), report that people typically disclose a WTP of approximately \$20.

The St. Petersburg Paradox was the earliest example of an 'anomaly' in choice behavior that led to a change in theory, insofar as it caused Bernoulli to supplant the Pascal–Fermat theory of Expected Monetary Value (EMV) with what has become known as Expected Utility (EU). Numerous alternative solutions to the paradox have subsequently been proposed.[12] Moreover, Yaari's (1987) dual theory of choice under risk has shown that the concave utility function (distortion of outcomes) solution is observationally indistinguishable from a distortion of probabilities solution, and that as such, concave utility is therefore not a necessary precondition for solving the St. Petersburg Paradox. Yet ultimately it was the mounting evidence of experimentally demonstrated EU-violating 'anomalies' – heuristics and biases of choice under risk and uncertainty – that allowed CPT to emerge as an alternative to EU. Although CPT serves as a descriptive model for a number of distinct behavioral biases and effects, alternation bias is not among them. Nevertheless, alternation bias is particularly relevant in the context of coin tossing sequences.

The next section shows that alternation bias is sufficient on its own to induce finite and moderate WTP for the St. Petersburg gamble. The following two sections show in turn that by controlling for alternation bias, currently popular CPT parameterizations do in fact satisfy appropriately specified finiteness constraints for the St. Petersburg gamble, and moreover they yield Certainty Equivalents (CEs) and WTP within the accepted empirical range.

3.1 Mathematical Expectation Revisited

In the present context, alternation bias enters the formulation of subjective (mathematical) expectation by distorting the subject's perception of the probability distribution of $\tilde{n}$, the index of the first toss on which an unbiased memoryless coin turns up 'Heads'.

Objectively $\tilde{n}$ follows a geometric distribution with parameter $p = \frac{1}{2}$, i.e. the objective probabilities are simply $p_n = \frac{1}{2}(1-\frac{1}{2})^{n-1} = 2^{-n}$ for $n = 1, 2, \ldots$.

Accounting for first-order alternation bias, under which $P(H|T) = 0.6$ and $P(T|T) = 0.4$, the subjectively perceived probability of the coin turning up 'Heads' for the first time on toss n then takes the form

$$p_n^{f\text{-}o} = \begin{cases} P(H) = \frac{1}{2} & \text{for } n = 1 \\ \frac{1}{2}P(H|T)P(T|T)^{n-2} = 0.3 \cdot 0.4^{n-2} & \text{for } n \geq 2 \end{cases}, \tag{1}$$

which gives a subjectively distorted mathematical expectation of

$$\begin{aligned} E^{f\text{-}o}(G_{StP}) &= \sum_{n=1}^{\infty} p_n^{f\text{-}o} 2^n = 1 + 0.3 \sum_{n=2}^{\infty} 0.4^{n-2} 2^n = 1 + \left(\frac{0.3}{0.4^2}\right) \sum_{n=2}^{\infty} 0.8^n \\ &= 1 + \left(\frac{0.3}{0.4^2}\right)\left[\frac{0.8}{0.2} - 0.8\right] = 7.0\,. \end{aligned} \tag{2}$$

So under first-order alternation bias alone, WTP is limited to \$7.0.

Accounting for higher-order alternation bias (see Table 1), the subjectively perceived probability of the coin turning up 'Heads' for the first time on toss n then takes the form

$$p_n^{h\text{-}o} = \begin{cases} P(H) = \frac{1}{2} & n = 1 \\ \frac{1}{2}P(H|T) = \frac{1}{2} \cdot 0.58 = 0.29 & n = 2 \\ \frac{1}{2}P(T|T)P(H|TT) = \frac{1}{2} \cdot 0.42 \cdot 0.62 = 0.1302 & n = 3 \\ \frac{1}{2}P(T|T)P(T|TT)P(H|TTT)P(T|TTT)^{n-4} = \frac{1}{2} \cdot 0.42 \cdot 0.38 \cdot 0.7 \cdot 0.3^{n-4} & \\ \qquad = 0.05586 \cdot 0.3^{n-4} & n \geq 4 \end{cases} \tag{3}$$

which gives a subjectively distorted mathematical expectation of

$$\begin{aligned} E^{h\text{-}o}(G_{StP}) &= \sum_{n=1}^{3} p_n^{h\text{-}o} 2^n + \sum_{n=4}^{\infty} p_n^{h\text{-}o} 2^n = \sum_{n=1}^{3} p_n^{h\text{-}o} 2^n + \left(\frac{0.05586}{3^4}\right) \sum_{n=4}^{\infty} 0.6^n \\ &= \sum_{n=1}^{3} p_n^{h\text{-}o} 2^n + \left(\frac{0.05586}{3^4}\right)\left[\frac{0.6^4}{0.4}\right] = 5.436\,. \end{aligned} \tag{4}$$

So under third-order alternation bias alone, WTP is limited to \$5.436.

Relative to the objective geometric distribution, first-order alternation bias induces a higher perceived probability of the coin-tossing sequence terminating on the second throw ($p_2 = 0.25 < 0.3 = p_2^{f\text{-}o}$) than on subsequent throws ($p_n > p_n^{f\text{-}o}\ \forall n \geq 3$), while third-order alternation bias induces a higher perceived probability of the coin-tossing sequence terminating on the second and third throws ($p_2 = 0.25 < 0.29 = p_2^{h\text{-}o}$, $p_3 = 0.125 < 0.1302 = p_3^{h\text{-}o}$) than on subsequent throws ($p_n > p_n^{h\text{-}o}\ \forall n \geq 4$).

Both the first-order (2) and higher-order (4) estimates of WTP induced by alternation bias alone fall within the generally accepted empirical range. Whereas Camerer (2005) shows that the 'anomalies literature' – through loss aversion[13] in particular – provides a solution to the St. Petersburg Paradox that requires neither a nonlinear value function nor a nonlinear (unconditional) probability weighting function, (2) and (4) show that the anomalies literature also gives rise to a second solution – based on alternation bias – which similarly makes no requirement for a nonlinear value function or a nonlinear (unconditional) probability weighting function.

3.2 CPT Finiteness Constraint Revisited

Blavatskyy (2005) and Rieger and Wang (2006) contend that conventional parameterizations of CPT fail to yield finite valuations for the St. Petersburg gamble, and that in order to resolve the St. Petersburg Paradox the parameterization of CPT must satisfy an additional constraint, namely $\alpha < \gamma$. Yet given what has been established above about the alternation bias – i.e. (a) that it is well-documented, (b) that it affects sequence trials exemplified by coin-tossing sequences such as those found in St. Petersburg gambles, and (c) that it has been established independently of unconditional probability distortion and non-linear outcome weighting – and given that CPT has been conceived as a descriptive theory to explain numerous heuristics and biases in choice under risk and uncertainty but *exclusive* of local representativeness effects, it is indeed no surprise at all that direct application of CPT to the St. Petersburg gamble proves problematic. For these very same reasons, however, it is neither necessary nor desirable to enforce the constraint $\alpha < \gamma$ even for the sole purpose of analyzing the St. Petersburg gamble. This holds with even more force for the parameterization of CPT for general use.

Incorporation of alternation bias into the analysis of the St. Petersburg gamble proceeds by way of distortion of conditional probabilities between coin tosses. Just as a casino player under the influence of the Gambler's Fallacy believes that his probability of winning this hand is higher *because* of a long sequence of losing hands leading up to this hand, an individual contemplating the St. Petersburg gamble under the influence of alternation bias believes that the probability of a particular toss turning up 'Heads' is higher *because* of an unbroken string of preceding 'Tails'.

As a consequence, the following two propositions may be proved using the estimates for first-order and higher-order alternation bias set out in Table 1. Proofs are collected in Appendix 1.

Proposition 1 (First-order constraint). *Once first-order alternation bias is controlled for, the finiteness constraint relaxes to*

$$\alpha < \frac{\log(5/2)}{\log(2)} \cdot \gamma \approx 1.32 \cdot \gamma \; . \tag{5}$$

Proposition 2 (Higher-order constraint). *Once alternation bias effects up to third order are controlled for, the finiteness constraint relaxes to*

$$\alpha < -\frac{\log(0.3)}{\log(2)} \cdot \gamma \approx 1.737 \cdot \gamma \; . \tag{6}$$

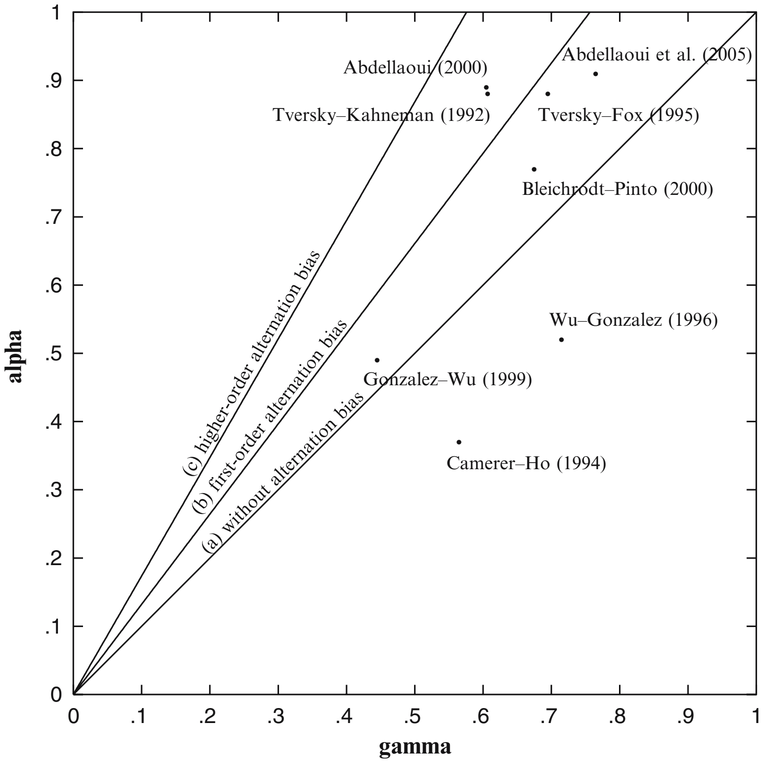


Fig. 1 Conventional parameterizations of cumulative prospect theory and the finiteness constraint computed (**a**) without alternation bias, (**b**) with first-order alternation bias, and (**c**) with higher-order alternation bias

As Fig. 1 illustrates, popular conventional parameterizations of CPT comfortably satisfy the finiteness constraint once it is adjusted for alternation bias up to third order.

Mathematically, the Blavatskyy–Rieger–Wang constraint without alternation bias, illustrated as (a) in Fig. 1, is derived from the limit behavior of the gross payout from the St. Petersburg gamble (G_{StP}). Similarly, the constraints (5) and (6) above, illustrated as (b) and (c) in Fig. 1, are also derived from the limit behavior of the gross payout from the St. Petersburg gamble (G_{StP}). Using numerical procedures it is possible to determine the Certainty Equivalent (CE) of this gross payout for each parameterization of CPT and for each of the three assumptions about alternation bias. The results of this numerical implementation are presented below in Table 2.

For the single-parameter probability weighting function specification, higher-order alternation bias brings the CE of the gross payout down to within the range $[5.64, 20.39]$. The distance of the parameter pair (γ, α) from the finiteness constraint is one determinant of the magnitude of this CE, but so is its location along the length of the finiteness constraint. Figures 2 and 3 illustrate this by way of the CE=\$5,

Table 2 Certainty equivalents of the gross payout from the St. Petersburg gamble under conventional parameterizations of CPT computed (**a**) Without alternation bias, (**b**) With first-order alternation bias, and (**c**) With higher-order alternation bias

		(γ,α)[14] (δ,γ,α)[15]	(a) Without alternation bias	(b) First-order alternation bias	(c) Higher-order alternation bias
Wu and Gonzalez (1996)					
	1-param $w^+(p)$	(0.71, 0.52)	16.00	6.95	5.64
	2-param $w^+(p)$	(0.84, 0.68, 0.52)	17.18	7.00	5.58
Camerer and Ho (1994)					
	1-param $w^+(p)$	(0.56, 0.37)	19.32	7.98	6.07
Abdellaoui et al. (2005)					
	1-param $w^+(p)$	(0.76, 0.91)	∞	22.08	8.18
	2-param $w^+(p)$	(0.98, 0.83, 0.91)	∞	11.39	6.75
Bleichrodt and Pinto (2000)					
	1-param $w^+(p)$	(0.67, 0.77)	∞	22.21	8.72
	2-param $w^+(p)$	(0.82, 0.55, 0.77)	∞	∞	16.23
Gonzalez and Wu (1999)					
	1-param $w^+(p)$	(0.44, 0.49)	∞	56.53	13.05
	2-param $w^+(p)$	(0.77, 0.44, 0.49)	∞	46.00	12.30
Tversky and Fox (1995)					
	1-param $w^+(p)$	(0.69, 0.88)	∞	74.81	10.15
	2-param $w^+(p)$	(0.76, 0.69, 0.88)	∞	56.15	8.53
Tversky and Kahneman (1992)					
	1-param $w^+(p)$	(0.61, 0.88)	∞	∞	17.39
Abdellaoui (2000)					
	1-param $w^+(p)$	(0.60, 0.89)	∞	∞	20.39

CE=\$10 and CE=\$20 contours for the gross payoff under first-order alternation bias and higher-order alternation bias respectively. The differences between these contour maps explain for instance why the Tversky and Fox (1995) parameterization yields a larger CE than the Gonzalez and Wu (1999) parameterization under first-order alternation bias $(74.81 > 56.53)$ while the reverse is true under higher-order alternation bias $(10.15 < 13.05)$.

3.3 WTP Under CPT Revisited

Nevertheless the above gross payout CE calculations should not be confused with WTP for the St. Petersburg gamble under CPT. Correct calculation of WTP under CPT must incorporate loss aversion over the shortfall between the gross payout G_{StP} and the up-front payment P exacted as the entry fee for participation in the St. Petersburg coin-tossing gamble. As Camerer (2005) points out,[16] attention must be focused on the net gamble payout $G_{StP} - P$, which involves an ex ante probable loss for $P > 2$. For each parameterization the maximum WTP will be less than the CE of the gross payout. Thus for any entry fee $P > 2$, the CPT evaluation occurs with respect to both gains and losses

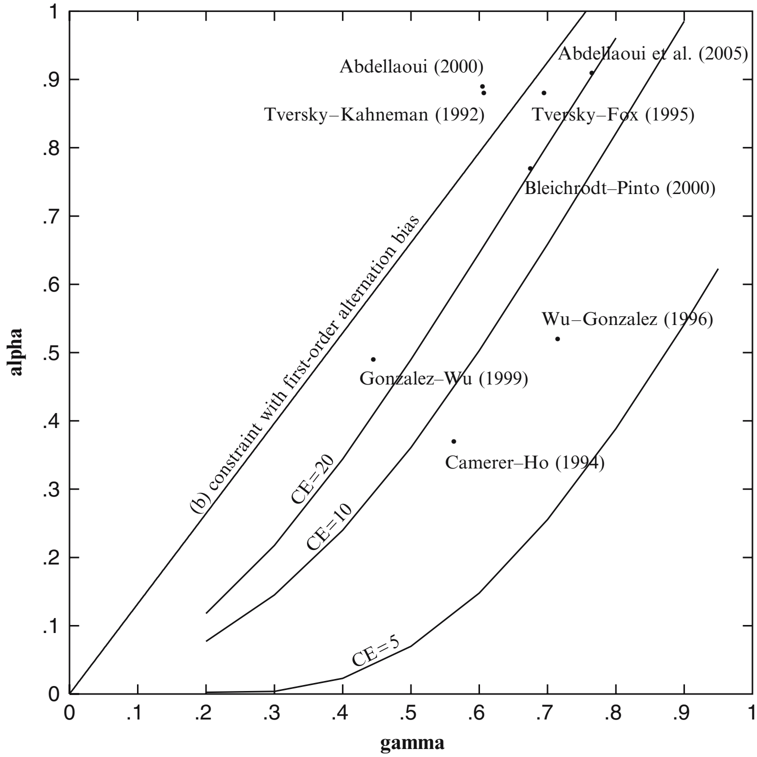


Fig. 2 The \$5, \$10, and \$20 Certainty Equivalent contours of the gross St. Petersburg gamble payout under first-order alternation bias

$$V(G_{StP}-P) = V^{+}((G_{StP}-P)^{+}) + V^{-}((G_{StP}-P)^{-}) \tag{7}$$

where the '+' superscript refers to gains and the '−' superscript refers to losses. The maximum WTP is the entry fee P^* that solves

$$V^{+}((G_{StP}-P^*)^{+}) + V^{-}((G_{StP}-P^*)^{-}) = 0\,. \tag{8}$$

For the Tversky and Kahneman (1992) parameters $\gamma^{\text{gains}} = 0.61$, $\alpha^{\text{gains}} = 0.88$, $\lambda = 2.25$, $\gamma^{\text{loss}} = 0.69$, $\alpha^{\text{loss}} = 0.88$ and third-order alternation bias, (8) is solved by $P^* \approx$ \$9.95, which is a finite maximum WTP that resolves the St. Petersburg Paradox.

4 Conclusion

If it could be shown that alternation bias is not operative in the St. Petersburg gamble setting, then present results would not diminish the case for restricting CPT parameterization in accordance with the Blavatskyy–Rieger–Wang finiteness constraint.

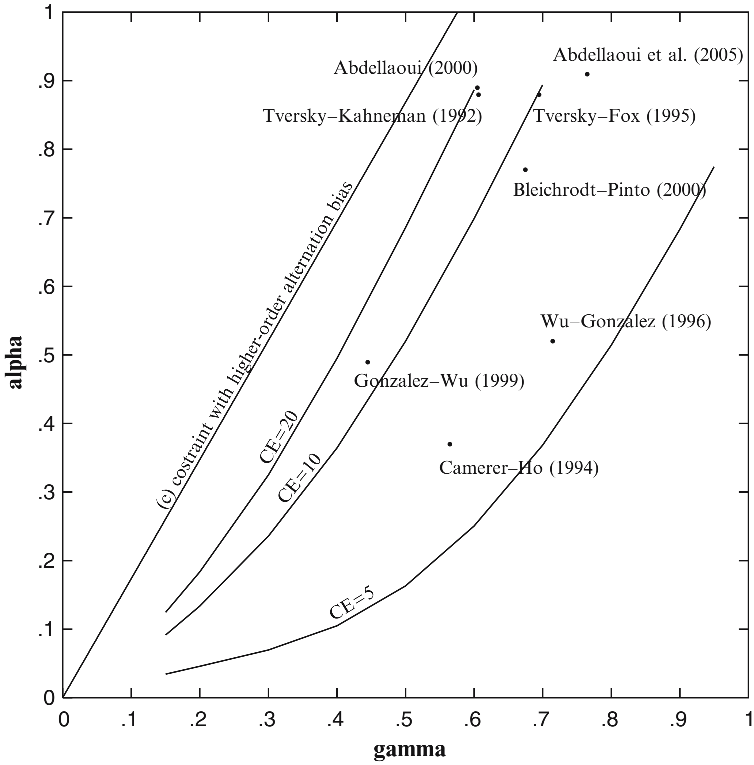


Fig. 3 The \$5, \$10, and \$20 Certainty Equivalent contours of the gross St. Petersburg gamble payout under higher-order alternation bias

However, experimental studies suggest strongly that coin-flipping series are indeed an exemplar contexts where alternation bias is operative and for which reliable and replicated empirical estimates of alternation bias magnitude are available.

The Blavatskyy (2005) and Rieger and Wang (2006) papers expose an important feature of conventional CPT parameterizations. Yet their elegantly straightforward remedy, namely the Blavatskyy–Rieger–Wang finiteness constraint ($\alpha < \gamma$) – although mathematically unobjectionable and certainly a solution worthy of consideration per se – is not without logical and theoretical consequences of its own. It localizes the remedy to the popular and conventional parameterizations of CPT, and these parameterizations are singled out as the effective cause of the finiteness problem. Yet the substantial experimental literature on the alternation bias suggests strongly that the solution to the finiteness problem may in fact lie in the subjective distortion of conditional probabilities, rather than in the subjective distortion of unconditional probabilities. The former (distortion of conditional probabilities), is formally outside the scope of CPT, whereas the latter (distortion of unconditional probabilities), is a proper part and object of the analytical structure of CPT.

To require CPT's unconditional probability distortion parameterization to reflect and incorporate the conditional probability distortion caused by alternation bias induced in the St. Petersburg gamble is to introduce a 'foreign' element into CPT (conditional probability distortion) and to do so in a way that confounds the magnitude of conditional probability distortion with the magnitude of unconditional probability distortion, as opposed to keeping the magnitudes of these two distinct effects separate and individually identifiable. Moreover, imposition of this constraint on parameterization limits the scope of applicability of CPT, insofar as the Blavatskyy–Rieger–Wang constraint rules out most of the widely used conventional parameterizations, which are tuned to achieving descriptive accuracy in a variety of settings that do not share the St. Petersburg gamble's sequential structure.

None of these concessions are necessary, though the cost of avoiding them is to bring more of the experimental and behavioral literature into the foreground. Recognizing the role of alternation bias in the St. Petersburg coin-tossing sequence allows the Paradox to be resolved, while preserving the distinction between conditional and unconditional probability distortion, and moreover preserving CPT's scope of descriptive applicability that is embodied in its conventional parameterizations.

As CPT becomes increasingly popular and is adopted and applied ever more widely, the question that is at the root of the divergence between the approach of this paper and that of Blavatskyy (2005) and Rieger and Wang (2006) will re-emerge with increasing frequency: How are we to apply, interpret and evaluate CPT? Is CPT a self-contained portable module that can be applied across the whole spectrum of problem settings without any need to anticipate complications, or is CPT essentially inseparable from the wider 'heuristics and biases' program? The special application studied in this paper lends weight to the latter. Although CPT has a concise, self-contained mathematical form, it should not be applied without giving due care and attention to the full range of behavioral effects that may arise. Some of these effects are captured by CPT, yet others require separate accommodation.

Appendix 1: Mathematical Appendix

The St. Petersburg gamble pays out $2^{\tilde{n}}$ where $\tilde{n} \in \mathbb{N}$ is the index of the first toss on which an unbiased memoryless coin turns up 'Heads'. Alternation bias alters the subjective perception of the distribution of $\tilde{n}$.

An individual with CPT preferences evaluates the gross St. Petersburg gamble payout G_{StP} as

$$V^{+}(G_{StP}) = \sum_{n=1}^{\infty} u^{+}(2^{n}) \cdot \left[w^{+}\left(\sum_{i=n}^{\infty} p_i\right) - w^{+}\left(\sum_{i=n+1}^{\infty} p_i\right)\right] \tag{9}$$

where

$$u^{+}(x) \equiv x^{\alpha} \qquad x \geq 0,\, \alpha \in (0,1) \tag{10}$$

is the value function for gains ($x \geq 0$) and

$$w^+(p) \equiv \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \qquad \gamma \in (0,1),\, p \in [0,1] \tag{11}$$

is the Tversky and Kahneman (1992) inverse S-shaped probability weighting function for gains. Blavatskyy (2005) shows that as $n \to \infty$ the denominator of the probability weighting function $w^+(p)$ converges to 1, and as attention may be restricted to the limit tail behavior, the approximation $w^+(p) \approx p^\gamma$ is valid, and thus in the case computed *without* alternation bias (9) simplifies to

$$V_w^+(G_{StP}) = (2^\gamma - 1)\sum_{n=1}^{\infty} 2^{(\alpha-\gamma)n} \ . \tag{12}$$

In order to ensure that the geometric series in (12) is convergent so that $V_w^+(G_{StP})$ remains finite, the constraint $\alpha < \gamma$ must be imposed (see finiteness constraint (a) in Fig. 1). However as Fig. 1 illustrates, this constraint is violated by most conventional parameterizations of CPT.

Finiteness Constraint with First-Order Alternation Bias

Consider the form of (9) with the first-order alternation bias that Kareev (1995) reports as being a standard finding in the literature: $P(H|T) = 0.6$ and $P(T|T) = 0.4$. The probability of the coin turning up 'Heads' for the first time on toss n then takes the form

$$p_n^{f\text{-}o} = \begin{cases} P(H) \ = \ \frac{1}{2} & \text{for } n = 1 \\ \frac{1}{2}P(H|T)P(T|T)^{n-2} \ = \ 0.3 \cdot 0.4^{n-2} & \text{for } n \geq 2 \end{cases} \tag{13}$$

and $\sum_{n=1}^{\infty} p_n^{f\text{-}o} = 0.5 + 0.3\sum_{j=0}^{\infty} 0.4^j = 0.5 + (0.3/0.6) = 1$. Therefore the first term in the outside sum of (9) is

$$a_1^{f\text{-}o} \ = \ u^+(2^1)\left[w^+\left(\sum_{i=1}^{\infty} p_i^{f\text{-}o}\right) - w^+\left(\sum_{i=2}^{\infty} p_i^{f\text{-}o}\right)\right] \ = \ u^+(2^1)[1 - w^+(2^{-1})] \ = \ 2^\alpha[1 - 2^{-\gamma}] \tag{14}$$

and subsequent terms are of the form

$$a_n^{f\text{-}o} = u^+(2^n)\left[w^+\left(\left(\frac{0.3}{0.4^2 0.6}\right)0.4^n\right) - w^+\left(\left(\frac{0.3}{0.4^2 0.6}\right)0.4^{n+1}\right)\right] \quad \forall n \geq 2 \tag{15}$$

$$= \left(\frac{0.3}{0.4^2 0.6}\right)^\gamma 2^{\alpha n}[0.4^{\gamma n} - 0.4^{\gamma(n+1)}] \qquad \forall n \geq 2 \tag{16}$$

giving a CPT evaluation of the gross St. Petersburg gamble payout G_{StP} under first-order alternation bias of

$$V^{+}_{f\text{-}o}(G_{StP}) = a_1^{f\text{-}o} + \sum_{n=2}^{\infty}\left(\frac{0.3}{0.4^2 0.6}\right)^{\gamma} 2^{\alpha n}[0.4^{\gamma n} - 0.4^{\gamma(n+1)}] \tag{17}$$

$$= a_1^{f\text{-}o} + \left(\frac{0.3}{0.4^2 0.6}\right)^{\gamma}\left[\sum_{n=2}^{\infty}\frac{2^{(\alpha+\gamma)n}}{5^{\gamma n}} - \left(\frac{2}{5}\right)^{\gamma}\sum_{n=2}^{\infty}\frac{2^{(\alpha+\gamma)n}}{5^{\gamma n}}\right] \tag{18}$$

$$= a_1^{f\text{-}o} + \left(\frac{0.3}{0.4^2 0.6}\right)^{\gamma}\left(1 - \frac{2^{\gamma}}{5^{\gamma}}\right)\sum_{n=2}^{\infty}\left(\frac{2^{(\alpha+\gamma)}}{5^{\gamma}}\right)^{n} \tag{19}$$

which is finite if the parameterization satisfies the constraint

$$\frac{2^{\alpha+\gamma}}{5^{\gamma}} < 1 \tag{20}$$

$$\alpha < \frac{\log(5/2)}{\log(2)}\cdot\gamma \quad \approx \quad 1.32\cdot\gamma\ . \tag{21}$$

This result is formalized as Proposition 1 and illustrated as finiteness constraint (b) in Fig. 1.

Finiteness Constraint with Higher-Order Alternation Bias

Using the transition probabilities up to third order presented in Table 1, the probability of the coin turning up 'Heads' for the first time on toss n then takes the form

$$p_n^{h\text{-}o} = \begin{cases} P(H) = \frac{1}{2} & n=1 \\ \frac{1}{2}P(H|T) = \frac{1}{2}\cdot 0.58 = 0.29 & n=2 \\ \frac{1}{2}P(T|T)P(H|TT) = \frac{1}{2}\cdot 0.42\cdot 0.62 = 0.1302 & n=3 \\ \frac{1}{2}P(T|T)P(T|TT)P(H|TTT)P(T|TTT)^{n-4} = \frac{1}{2}\cdot 0.42\cdot 0.38\cdot 0.7\cdot 0.3^{n-4} \\ \qquad = 0.05586\cdot 0.3^{n-4} & n\geq 4 \end{cases} \tag{22}$$

and $\sum_{n=1}^{\infty} p_n^{h\text{-}o} = 0.5 + 0.29 + 0.1302 + 0.05586\sum_{j=0}^{\infty} 0.3^j = 0.9202 + \frac{0.05586}{0.7} = 1$. The first, second and third terms in the outside sum of (9) are

$$a_1^{h\text{-}o} = u^{+}(2^1)\left[w^{+}\left(\sum_{i=1}^{\infty} p_i^{h\text{-}o}\right) - w^{+}\left(\sum_{i=2}^{\infty} p_i^{h\text{-}o}\right)\right] = u^{+}(2^1)[1 - w^{+}(2^{-1})] = 2^{\alpha}[1 - 2^{-\gamma}]\ , \tag{23}$$

$$a_2^{h\text{-}o} = u^{+}(2^2)\left[w^{+}\left(\sum_{i=2}^{\infty} p_i^{h\text{-}o}\right) - w^{+}\left(\sum_{i=3}^{\infty} p_i^{h\text{-}o}\right)\right] = u^{+}(2^2)[w^{+}(2^{-1}) - w^{+}(1 - 0.5 - 0.29)]$$

$$= 2^{2\alpha}[2^{-\gamma} - 0.21^{\gamma}]\ , \tag{24}$$

and

$$a_3^{h\text{-}o} = u^+(2^3)\left[w^+\left(\sum_{i=3}^{\infty} p_i^{h\text{-}o}\right) - w^+\left(\sum_{i=4}^{\infty} p_i^{h\text{-}o}\right)\right] = u^+(2^3)[w^+(0.21)$$
$$- w^+(1-0.5-0.29-0.1302)] = 2^{3\alpha}[0.21^\gamma - 0.0798^\gamma]\,. \tag{25}$$

Subsequent terms ($\forall n \geq 4$) are of the form

$$a_n^{h\text{-}o} = u^+(2^n)\left[w^+\left(\sum_{i=n}^{\infty} p_i^{h\text{-}o}\right) - w^+\left(\sum_{i=n+1}^{\infty} p_i^{h\text{-}o}\right)\right] \tag{26}$$

$$= u^+(2^n)\left[w^+\left(0.05586\sum_{i=n}^{\infty} 0.3^{i-4}\right) - w^+\left(0.05586\sum_{i=n+1}^{\infty} 0.3^{i-4}\right)\right] \tag{27}$$

$$= u^+(2^n)\left[w^+\left(\frac{0.05586}{0.3^4}\sum_{i=n}^{\infty} 0.3^{i}\right) - w^+\left(\frac{0.05586}{0.3^4}\sum_{i=n+1}^{\infty} 0.3^{i}\right)\right] \tag{28}$$

$$= u^+(2^n)\left[w^+\left(\frac{0.05586}{0.3^4\cdot 0.7}0.3^{n}\right) - w^+\left(\frac{0.05586}{0.3^4\cdot 0.7}0.3^{n+1}\right)\right] \tag{29}$$

$$= \left(\frac{0.05586}{0.3^4\cdot 0.7}\right)^\gamma 2^{\alpha n}[0.3^{\gamma n} - 0.3^{\gamma(n+1)}]\ . \tag{30}$$

Thus the CPT evaluation of the gross St. Petersburg gamble payout G_{StP} under higher-order alternation bias may be written as

$$V_{h\text{-}o}^+(G_{StP}) = \sum_{n=1}^{3} a_n^{h\text{-}o} + \sum_{n=4}^{\infty}\left(\frac{0.05586}{0.3^4\cdot 0.7}\right)^\gamma 2^{\alpha n}[0.3^{\gamma n} - 0.3^{\gamma(n+1)}] \tag{31}$$

$$= \sum_{n=1}^{3} a_n^{h\text{-}o} + \left(\frac{0.05586}{0.3^4\cdot 0.7}\right)^\gamma\left[\sum_{n=4}^{\infty}(2^\alpha 0.3^\gamma)^n - 0.3^\gamma\sum_{n=4}^{\infty}(2^\alpha 0.3^\gamma)^n\right] \tag{32}$$

$$= \sum_{n=1}^{3} a_n^{h\text{-}o} + \left(\frac{0.05586}{0.3^4\cdot 0.7}\right)^\gamma (1-0.3^\gamma)\sum_{n=4}^{\infty}(2^\alpha 0.3^\gamma)^n \tag{33}$$

which converges to a finite value if the parameterization satisfies the constraint

$$2^\alpha 0.3^\gamma < 1 \tag{34}$$

$$\alpha < -\frac{\log(0.3)}{\log(2)}\cdot\gamma \quad \approx \quad 1.737\cdot\gamma\ . \tag{35}$$

This result is formalized as Proposition 2 and illustrated as finiteness constraint (c) in Fig. 1.

Notes

[1]For example, Abdellaoui (2000), Abdellaoui et al. (2005), Bleichrodt and Pinto (2000), Gonzalez and Wu (1999), Tversky and Fox (1995), and Tversky and Kahneman (1992).

[2]The certainty effect; overweighting small probabilities, underweighting large probabilities.

[3]That is, the reflection effect; risk seeking in losses and risk aversion in gains; diminishing sensitivity, whereby individuals are more sensitive to changes near their status quo than to changes that are more remote from their status quo.

[4]Losses are weighed more heavily than gains.

[5]By not invoking or bundling monetary payoffs with 'Heads' or 'Tails' realizations.

[6]See e.g. Clotfelter and Cook (1993), Terrel (1994) and Croson and Sundali (2005).

[7]Also known as the 'three door problem'; it is mathematically equivalent to the 'three prisoner problem'. Although bias is pervasive in these problems (Granberg & Brown, 1995; Granberg, 1999), nevertheless it is possible to devise schemes that allow subjects to learn how to overcome their anomalous initial biases (Friedman, 1998; Krauss & Wang, 2003). In a market setting, the presence of a small proportion of bias-free agents suffices to eliminate bias in prices (Kluger & Wyatt, 2004).

[8]In Reichenbach's terminology, a 'negative recency' effect.

[9]Or the 'law of small numbers'.

[10]In Bernoulli's (1738) variant of the St. Petersburg gamble, the subject's (Paul's) payout is $2^{\tilde{n}-1}$ ducats.

[11]If the payout is specified as $2^{\tilde{n}-1}$ dollars, then the certainty equivalent associated with the logarithmic utility function is 2 currency units: $E[u(2^{\tilde{n}-1})] = \sum_{n=1}^{\infty} 2^{-n}\log(2^{n-1}) = \sum_{n=1}^{\infty} \frac{n-1}{2^n}\log(2) = \log(2)$.

[12]Alternative solutions proposed for the St. Petersburg Paradox are too numerous to be discussed in detail here. For reviews see Samuelson (1977), Vlek and Wagenaar (1979) and Bottom et al. (1989).

[13]In conjunction with piecewise linear utility and large but finite upper ceiling on the maximum payout.

[14]$w^+(p) = p^\gamma/(p^\gamma + (1-p)^\gamma)^{1/\gamma}$, $u^+(x) = x^\alpha$.

[15]$w^+(p) = \delta p^\gamma/(\delta p^\gamma + (1-p)^\gamma)$, $u^+(x) = x^\alpha$.

[16]Camerer (2005) uses piecewise linear utility, loss aversion, and the realistic assumption of the existence of a finite maximum payout ceiling to show that risk aversion is not a necessary condition for resolution of the St. Petersburg Paradox.

Acknowledgements I wish to thank David Peel, Bob Rothschild, Horst Zank, Pavlo Blavatskyy, David Budescu, John Hey, along with session participants at the Rome 2006 FUR XII meeting, the Nottingham 2006 ESA European meeting and the BGU/UCLA 2006 Affect, Motivation, and Decision Making conference, for their comments and advice, as well as William Bottom, David Budescu (a second time) and Colin Camerer for helpfully responding to my queries concerning the experimental and empirical literature. The usual disclaimer applies.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utilities and probability weighting functions. *Management Science, 46*(11), 1497–1512

Abdellaoui, M., Vossmann, F., & Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science, 51*(9), 1384–1399

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics, 12*(4), 428–454

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis, Commentarii Academiae Scientiarum Imperialis Petropolitanae

Blavatskyy, P. R. (2005). Back to the St. Petersburg paradox? *Management Science, 51*(4), 677–678

Bleichrodt, H., & Pinto, J. M. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science, 46*(11), 1485–1496

Bottom, W. P., Bontempo, R. N., & Holtgrave, D. R. (1989). Experts, novices, and the St. Petersburg Paradox: Is one solution enough? *Journal of Behavioral Decision Making, 2*(3), 139–147

Budescu, D. V. (1987). A markov model for generation of random binary sequences. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 25–39

Camerer, C. (2005). Three cheers – psychological, theoretical, empirical – for loss aversion. *Journal of Marketing Research, 42*(2), 129–133

Camerer, C., & Ho., T.- H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty, 8*(2), 167–196

Chernoff, H., & Moses, L. (1959). *Elementary decision theory.* New York, NY: Wiley

Clotfelter, C., & Cook, P. (1993). The "Gambler's Fallacy" in lottery play. *Management Science, 39*(12), 1521–1525

Croson, R., & Sundali, J. (2005). The Gambler's Fallacy and the Hot Hand: Empirical data from casinos. *Journal of Risk and Uncertainty, 30*(3), 195–209

Friedman, D. (1998). Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *American Economic Review, 88*(4), 933–946

Gilovich, T., Vallone, R., & Tversky, A. (1985). The Hot Hand in basketball: On the misperception of random sequences. *Cognitive Psycology, 17*(3), 295–314

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*(1), 129–166

Granberg, D. (1999). Cross-cultural comparison of responses to the Monty Hall dilemma. *Social Behavior and Personality, 27*(4), 431–438

Granberg, D., & Brown, T. A. (1995). The Monty Hall dilemma. *Personality and Social Psychology Bulletin, 21*(7), 711–723

Kareev, Y. (1995). Positive bias in the perception of covariation. *Psychological Review, 102*(3), 490–502

Kluger, B. D., & Wyatt, S. B. (2004). Are judgment errors reflected in market prices and allocations? Experimental evidence based on the Monty Hall problem. *Journal of Finance, 59*(3), 969–997

Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall problem: Discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General, 132*(1), 3–22

Rabin, M. (2002). Inference by believers in the law of small numbers. *Quarterly Journal of Economics, 117*(3), 775–816

Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behavior. *Psychological Review, 104*(3), 603–617

Reichenbach, H. (1934). *Wahrscheinlichkeitslehre: Eine Untersuchung uber die logischen und mathematischen Grundlagen der Wahrscheinlichkeitsrechnung*. Translated as Reichenbach, H. (1949). *The theory of probability; An inquiry into the logical and mathematical foundations of the calculus of probability*. Berkeley, CA: University of California Press

Rieger, M. O., & Wang, M. (2006). Cumulative prospect theory and the St. Petersburg Paradox. *Economic Theory, 28*(3), 665–679

Samuelson, P. A. (1977). St. Petersburg Paradoxes: Defangled, dissected, and historically described. *Journal of Economic Literature, 15*(1), 24–55

Schmeidler, D., & Wakker, P. (1998). Expected utility and mathematical expectation. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The new Palgrave; A dictionary of economics* (pp. 229–232). London: Macmillan

Terrel, D. (1994). A test of the Gambler's Fallacy: Evidence from pari-mutuel games. *Journal of Risk and Uncertanty, 8*(3), 309–317

Tversky, A., & Fox, C. (1995). Weighing risk and uncertainty. *Psychological Review, 102*(2), 269–283

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1130

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323

Vlek, C., & Wagenaar, W. A. (1979). Judgement and decision under uncertainty. In J. A. Michon, E. G. J. Eijkman, & L. F. W. de Klerk (Eds.), *Handbook of psychonomics – Vol. 2*. Amsterdam: North-Holland

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science, 42*(12), 1676–1690

Yaari, M. (1987). The dual theory of choice under risk. *Econometrica, 55*(1), 95–115

Proposing a Normative Basis for the S-Shaped Value Function

M.E. Fabiyi

Keywords: S-shaped value function · Prospect theory · Evolutionary theory · Fitness

1 Introduction

The S-shaped value function is pervasive in its occurrence in preference tasks and prospect theory is based on the view that all preference tasks are comprehensively captured by an S-shaped value function which is concave in gains and convex for losses. In this work, we propose a normative basis for the ubiquitous S-shaped value function of Kahneman and Tversky's prospect theory (1979, 1992). We argue that it is best accommodated within the paradigmatic context of evolutionary theory. Proceeding from this paradigm, we show that the S-shaped value or utility function must necessarily result from rational actions by the decision agent.

The value function, $v(x)$ for changes in wealth is normally concave above the reference point i.e., $v''(x) < 0$, for $x > 0$ and often convex below it i.e., $v''(x) > 0$, for $x < 0$ (Kahneman & Tversky, 1979). These bounds imply that the value function is S-shaped. Kahneman and Tversky have summarized the properties of the S-Shaped value function as being: (a) generally concave for gains and convex for losses; (b) steeper for losses than for gains, and (c) defined on deviations from the reference point.

Although the pervasive occurrence of the S-shaped value function has been confirmed in the field (Collins, Musser, & Mason, 1991; Fishburn & Kochenberger, 1979; Galanter & Pliner, 1974), there is as yet no satisfactory normative explanation

M.E. Fabiyi
University of Chicago Graduate School of Business, The University of Chicago Graduate School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA
e-mail: mfabiyi@uchicago.edu

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.

© Springer-Verlag Berlin Heidelberg 2008

as regards why it should be so shaped. There have of course been prior attempts at providing a rationale for the presence of S-shaped value functions. For example, Friedman (1989) has suggested that an S-shaped value function emerges when agents maximize expected sensitivity at actual choice opportunities. Friedman's approach requires that the agents are capable of making *plausible* assumptions about the distribution of such opportunities. This requirement that the decision agent have some prior knowledge of opportunity distributions does not seem to us consistent with the reality that oftentimes, decision agents do not possess such information. Robson (2002) has argued that a utility function emerges from natural selection considerations framed as a principal agent problem – where nature is the principal.

This inability to ascribe a normative basis to the S-shaped value function amongst other things, is probably responsible for the descriptive characterization of prospect theory (Tversky & Kahneman, 1992). Yet, the ubiquitous emergence of the S-shaped utility function in preference elicitation tasks suggests that the properties that result in such a curve are intrinsic to the human species and probably to all organisms that are fitness maximizers.

The following fundamental questions must bear on the minds of researchers in the field of rational choice, these being: (a) Why is the utility curve S shaped? (b) Why is the curve steeper in losses than it is in gains? (c) What is the normative basis for the S-shaped utility profile? (d) What factors determine the shape of the curve and how can they be influenced? (e) Can the utility curve be established a priori? (f) Is the utility curve valid across temporal dimensions? (g) What is the significance of the reference point, how is it to be located, and can it be identified a priori?

2 An Evolutionary Rationale for the S-Shaped Value Function

As we have already indicated, attempts to use evolutionary theory to provide a rationale for the S-shaped character of the value function are not new. Rust (2004) derives a survival value function with an S-shaped profile from an evolutionary scenario in which behavioral tendencies naturally select with survival, according to a hazard model. In this model, the hazard rate requires decomposition into an internal and external (environmental) component and choices made by the decision agent result in adjustments to the external (environmental) hazard component. Although the survival value function derived by Rust differs from that of prospect theory significantly, as it has finite limits, it is nevertheless instructive that S-shaped value or utility functions can result from applying purely evolutionary considerations to the behavior of decision agents.

Ultimately, the evolutionary basis underlying the existence of a preference curve stems from the fact that every behavior manifested by a decision agent in response to a consumption stimulus has an impact on the capacity for survival. The behavior manifested by a decision agent in response to a stimulus (or prospect) results in either an enhancement of fitness, a reduction in fitness, or it might be fitness neutral i.e., possessing of a null effect on fitness. Given the importance of behavior for survival, it is suggested that psychological adaptations would have evolved which serve fundamentally to influence the decision making process in a manner that maximizes fitness. Proceeding from this premise, we posit the following:

1. The goal of all organisms is to achieve the perpetual representation of the organisms' genes in the gene pool
2. Organisms achieve this by birthing progeny, and working to assure the survival of the progeny until sexual maturity
3. The progenitor organism seeks at all times to maximize total fitness, i.e., the joint survival of the progeny (hereafter referred to as gene survival, G_s) and the progenitor (hereafter referred to as self survival, S_s)
4. The survival set is comprised of two elements, these being, gene survival and self survival

3 The Impact of Prospects and Behavior on Fitness

All stimuli or prospects serve to evince behavioral responses from the decision agent. Following from the assumption that the survival set is comprised of two elements – gene and self survival elements, we can conceive of nine (9) effects of behavior on the fitness set. These nine *goal states* emerge from the reality that a behavior can have any one of three effects on an entity, these being: (a) enhance fitness, (b) diminish fitness, or (c) have no impact on fitness. We assign positive (+), negative (−), and neutral (0) signs to these outcomes, respectively (see Table 1 for a listing of the goal states).

In order to establish the magnitude of the goal state at any one time, the values associated with the members of the fitness set are simply added up. To facilitate this, we need to make the following assumptions

1. Real number weights can be assigned to the members of the fitness set
2. The valence of the members of the fitness set carry nominal values given as +1 for the positive sign, −1 for the negative sign and 0 for the neutral (zero) sign
3. The magnitudes of the members of the fitness set are given as G, and S respectively for the gene and self survival members
4. A relative magnitude ϕ, of the weights of the members of the fitness set is defined, and is given as: $\frac{G}{S} = \phi$

Table 1 Goal or Motivational States

	G(+)	G(0)	G(−)
S(+)	G(+)S(+)	G(0)S(+)	G(−)S(+)
S(0)	G(+)S(0)	G(0)S(0)	G(−)S(0)
S(−)	G(+)S(−)	G(0)S(−)	G(−)S(−)

Note. The nine motivational or goal states. The sets obey commutative laws of addition thus $G(+)S(+) = S(+)G(+)$

Table 2 Derivation of fitness values from goal states. The normalized fitness values are obtained by dividing the fitness values by the maximum fitness value obtained across all nine goal states. For all $\frac{G}{S} \geq 1$, the goal state G(+)S(+) will always return the highest fitness value

Goal state	Fitness value	Normalized fitness value
G(+)S(+)	$S(\phi+1)$	$\frac{S(\phi+1)}{S(\phi+1)} = 1$
G(+)S(0)	$S(\phi)$	$\frac{S(\phi)}{S(\phi+1)} = \frac{\phi}{\phi+1}$
G(+)S(−)	$S(\phi-1)$	$\frac{S(\phi-1)}{S(\phi+1)} = \frac{\phi-1}{\phi+1}$
G(0)S(+)	S	$\frac{S}{S(\phi+1)} = \frac{1}{\phi+1}$
G(0)S(0)	0	$\frac{0}{S(\phi+1)} = 0$
G(0)S(−)	$-S$	$\frac{-S}{S(\phi+1)} = \frac{-1}{\phi+1}$
G(−)S(+)	$S(1-\phi)$	$\frac{S(1-\phi)}{S(\phi+1)} = \frac{1-\phi}{\phi+1}$
G(−)S(0)	$-S(\phi)$	$\frac{-S(\phi)}{S(\phi+1)} = \frac{-\phi}{\phi+1}$
G(−)S(−)	$-S(\phi+1)$	$\frac{-S(\phi+1)}{S(\phi+1)} = -1$

5. The relative magnitude of the weights of the members of the fitness set is specified within the bound $-\infty \leq \frac{G}{S} \leq \infty$. The maximization of fitness (assurance of perpetual presence of genes in the gene pool) requires that $\frac{G}{S} \geq 1$. This is so because gene permanence is a reality that is possible only if gene survival is assured.

Since we have defined $\frac{G}{S} = \phi$, then $G = S\phi$. We further define the value of the fitness set as F and for the goal state $G(\alpha)S(\beta)$, we have that:

$$F = G(\alpha) + S(\beta) \tag{1}$$

Substituting for G in the expression, we obtain

$$F = S\phi(\alpha) + S(\beta) = S(\phi\alpha + \beta) \tag{2}$$

Recalling that α and β can take any of the values +1, 0, and -1, we can derive real number fitness values for the various goal states (see Table 2).

4 Deriving the S-Shaped Value/Utility Function

We now proceed to develop a method for deriving utility from the goal states. From the purview of the approach adopted in this work, the ultimate impact or value of any stimulus is reducible to its effect on fitness (Robson, 2002). Thus, every stimulus and the behavior that it evinces, has a consequence on fitness – serving to increase or reduce fitness from some current value. That current value will be the reference state around which the consequence of the stimulus will be evaluated. Since we can determine the consequence of a stimulus from its effect on the goal state, then the *utility or value* of a stimulus is equivalent to the magnitude of its impact on survival. We posit therefore that the *motive or goal* states correspond to utility states since they are the currency by which the fitness impacts of stimuli are measured.

From an evolutionary perspective, the ultimate goal of the individual is the attainment of the G(+)S(+) state,[1] and at a minimum, evolutionary imperatives require the avoidance of the G(−) and, or S(−) states[2] since these latter states can potentially yield negative (hence minimizing) values for the fitness function. We can therefore conceive of a mapping function with which to transform the fitness values, F to their equivalent utilities, U, i.e., we seek the transform $\Re : F \rightarrow U$. By defining the normalized fitness value as x, we can specify the transform relation as:

$$\Re : \begin{cases} F^{+} \rightarrow U^{+} & x \geq 0 \\ F^{-} \rightarrow U^{-} & x < 0 \end{cases} \tag{3}$$

Given the wide range of stimuli that can be presented to the decision agent, it is best to utilize normalized quantities for the fitness values and their corresponding utilities. We define these as x and u respectively. Since the normalized fitness values are bounded within the range $-1 \leq x \leq 1$ we specify additionally that the normalized utility be bounded also within the domain $-1 \leq u \leq 1$.

In the positive domain the transform relation is unambiguous. Since the maximum attainable goal state G(+)S(+) corresponds to $x = +1$, the corresponding utility is $u = +1$. An optimized decision algorithm should indicate no utilities higher than would correspond to the goal state G(+)S(+), since none should exist. Whereas the transformation of fitness values to utilities on the positive dimension is intuitive, establishing how to conceive of the transformation from fitness value to utilities in the negative domain is not as obvious.

We assume that fitness maximization constraints require that no fitness utility is derivable from any goal state that compromises the *total survival* set i.e. when $x < 0$ then $u = -1$. Thus, the transform $x^{-} \rightarrow u^{-}$ is minimized not at G(−)S(−), but at any of the G(−), S(−) states that returns the first negative fitness value. Based on these considerations, we can specify that:

$$u(x) = \begin{cases} u_{\max} = +1 \Rightarrow G(+)S(+) & x \geq 0 \\ u_{\min} = -1 \Rightarrow Max\{\mathrm{G}(-)\mathrm{S}(-);\ \mathrm{G}(-)\mathrm{S}(0);\ \mathrm{G}(-)\mathrm{S}(+);\ \mathrm{G}(0)\mathrm{S}(-)\} & x < 0 \end{cases} \tag{4}$$

It is required that $u(x)$ is continuous, differentiable and bounded within the domain $u(x) : \Re \rightarrow (-1, 1)$. Specifying this constraint necessarily limits the choice of functional types suitable for the transformation. We will adopt the use of the hyperbolic Tan (*Tanh*) as a transform function that satisfies the necessary conditions.[3] We therefore have that $f(x) = Tanh(x)$ and we reformulate the problem thus:

$$u(x) = \begin{cases} Tanh(\alpha^{+}x) & x \geq 0 \\ Tanh(\alpha^{-}x) & x < 0 \end{cases} \tag{5}$$

Where α^{+} and α^{-} are constants, and generally $\alpha^{-} = \phi\alpha^{+}$. We now proceed to illustrate the method for $\phi = 3$ (see Tables 3 and 4). We first derive the value of α^{+}

Table 3 This table depicts the methodology for the derivation of the fitness function. For this example $\phi = 3.0$

	Fitness (F)		
	$G(+) = 3$	$G(0) = 0$	$G(-) = -3$
$S(+) = 1$	$= 3 + 1 = 4$	$= 0 + 1 = 1$	$= -3 + 1 = -2$
$S(0) = 0$	$= 3 + 0 = 3$	$= 0 + 0 = 0$	$= -3 + 0 = -3$
$S(-) = -1$	$= 3 + -1 = 2$	$= 0 + -1 = -1$	$= -3 + 1 = -4$

Table 4 Normalized fitness values for $\phi = 3.0$. Note that Table 4 is derived by normalizing Table 3

	Normalized fitness (x)		
	$G(+) = 3$	$G(0) = 0$	$G(-) = -3$
$S(+) = -1$	1	0.25	−0.5
$S(0) = 0$	0.75	0	−0.75
$S(-) = -1$	0.5	−0.25	−1

Table 5 Transform relations for the normalized fitness values for $\phi = 3.0$

	$\phi = 3$	
G-S States	x	u
G(+)S(+)	1.00	1.0
G(+)S(0)	0.75	1.0
G(0)S(+)	0.50	0.9
G(+)S(−)	0.25	0.6
G(0)S(0)	0.00	0.0
G(−)S(+)	−0.25	−1.0
G(0)S(−)	−0.50	−1.0
G(−)S(0)	−0.75	−1.0
G(−)S(−)	−1.00	−1.0

necessary to ensure that $u = (1, -1)$ at $x = (1, -1)$, and we obtain $\alpha^+ = 3$. Hence at this condition, we have that

$$u(x) = \begin{cases} Tanh(3x) & x \geq 0 \\ Tanh(3\phi x) & x < 0 \end{cases} \tag{6}$$

For this example, α^+ and α^- have the values 3 and 9. Hence, we have that for $x \geq 0$, $u(x) = Tanh(3x)$; and for $x < 0$, $u(x) = Tanh(9x)$. Using these relations, we obtain a utility curve with the characteristic S-shape (see Tables 3–5 and Fig. 1).

By applying the same approach, we can derive utility curves for the conditions specified by ϕ = 1, 9 and 0.25 (see Tables 6–8). Figures 1–4 depict the utility function profiles for $\phi = 3, 1, 9$ and 0.25 respectively. It is striking that all the profiles are generally S-shaped, although differences exist in the character of the S-shaped profiles that are derived.

Thus, we find that an evolutionary approach leads us, from a set of simple assumptions, to a utility function that accommodates the S shaped value function of

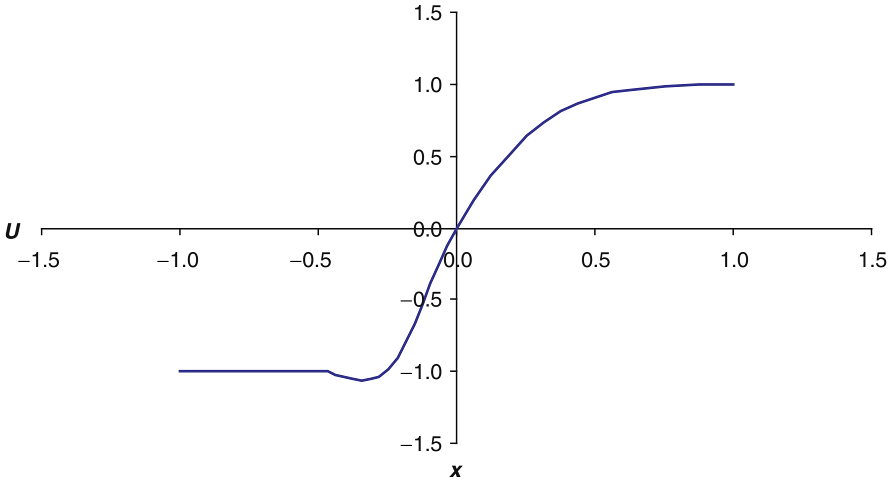


Fig. 1 Utility function derived from Table 5. $\phi = 3.0$, and α^+ and α^- have the values 3 and 9

Table 6 Transform relations for the normalized fitness values for $\phi = 1.0$

	$\phi = 1$	
G-S States	x	u
G(+)S(+)	1.0	1.0
G(+)S(0)	0.5	1.0
G(0)S(+)	0.5	1.0
G(+)S(−)	0.0	0.0
G(0)S(0)	0.0	0.0
G(−)S(+)	0.0	0.0
G(0)S(−)	−0.5	−1.0
G(−)S(0)	−0.5	−1.0
G(−)S(−)	−1.0	−1.0

Table 7 Transform relations for the normalized fitness values for $\phi = 9.0$

	$\phi = 9$	
G-S States	x	u
G(+)S(+)	1.0	1.0
G(+)S(0)	0.9	1.0
G(0)S(+)	0.8	1.0
G(+)S(−)	0.1	0.3
G(0)S(0)	0.0	0.0
G(−)S(+)	−0.1	−1.0
G(0)S(−)	−0.8	−1.0
G(−)S(0)	−0.9	−1.0
G(−)S(−)	−1.0	−1.0

Table 8 Transform relations for the normalized fitness values for $\phi = 0.25$

	$\phi = 0.25$	
G-S States	x	u
G(+)S(+)	1	1.00
G(0)S(+)	0.8	1.00
G(−)S(+)	0.6	1.00
G(+)S(0)	0.2	1.00
G(0)S(0)	0	0.00
G(−)S(0)	−0.2	−0.99
G(+)S(−)	−0.6	−1.00
G(0)S(−)	−0.8	−1.00
G(−)S(−)	−1	−1.00

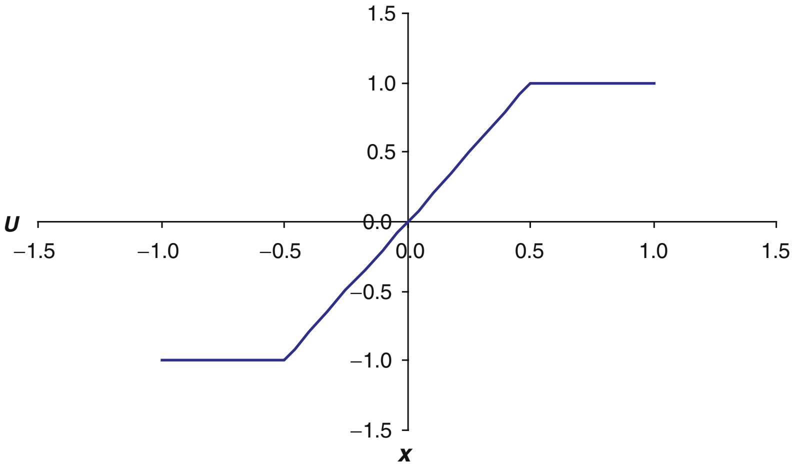


Fig. 2 Utility function derived from Table 6. $\phi = 1.0$, and α^+ and α^- both have the value 22.2

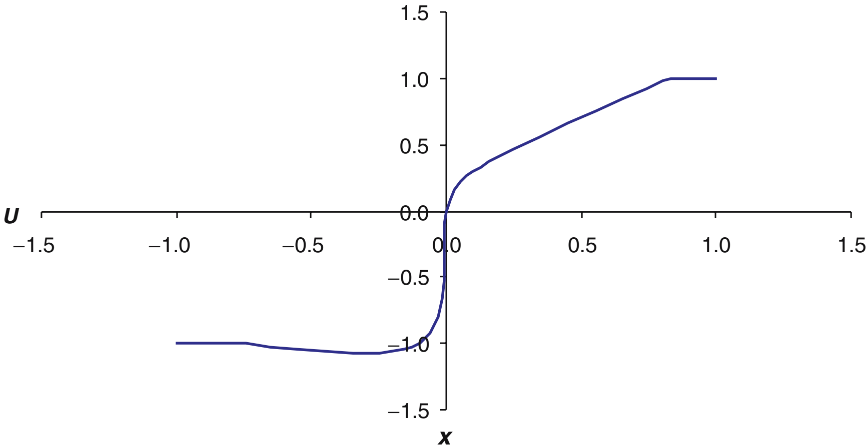


Fig. 3 Utility function derived from Table 7. $\phi = 9$, and α^+ and α^- are 3 and 27 respectively

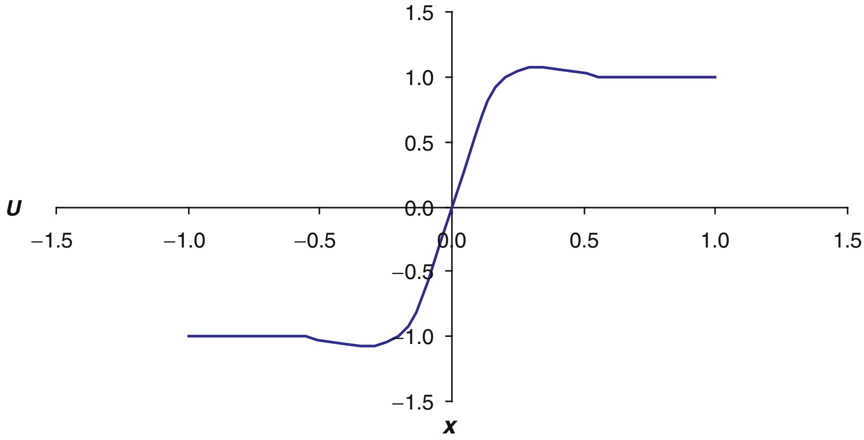


Fig. 4 Utility function derived from Table 8. $\phi = 0.25$, and α^+ and α^- have the values 53.2 and 13.3

prospect theory. We have suggested earlier that a necessary condition for gene permanence is that $\phi > 1$. Should this hold true, we will find that for all x, $\alpha^+ < \alpha^-$, implying that the value function will always be steeper in losses than in gains.[4]

5 Discussion

The utility function that we have derived demonstrates the properties of *diminishing sensitivity* and *loss aversion*. Evolutionary theoretic considerations therefore lead quite naturally to the specification for a utility function that is concave in gains and convex in losses. This treatment, as we have shown, accommodates the S shaped value function and provides a normative basis for its shape. That shape, we believe, is defined by the unique fitness weights, ϕ, of the individual. Our approach suggests that although the utility function of all gene fitness dominant individuals ($\phi > 1$) should be S-shaped, the precise slopes of the utility curve in either gains or losses will generally vary between individuals according to differences in the parameters ϕ and α^+. These parameters, we expect, can be readily determined by the estimation of a decision agent's value function using the method of certainty equivalents.

Our work strongly suggests that the S-shaped value function proceeds naturally from a behavioral framework that has as its objective function the maximization of survival.

This result has significant implications for choice tasks. First, it establishes the possibility for the a priori estimation of the value function for an individual. It is our view that the value function established for a decision agent will be reproduced in any elicitation task, so long as the parameters ϕ and α^+ remain constant. We conceive of these parameters as being invariant for any given individual. Finally, while we hypothesize that evolutionary imperatives would have selected for gene fitness maximizers ($\phi > 1$), the survival of strictly self fitness maximizers ($\phi < 1$) is not precluded.

Notes

[1]This is consistent with notions of the rational agent as a utility maximizer.

[2]These could be any of the states G(−)S(−), G(−)S(0), G(−)S(+) or G(0)S(−).

[3]Several variants of the sigmoid function could potentially also satisfy the necessary conditions (see Rojas, 1996).

[4]This will always be the case on condition that the weight of the G set is higher than that of the S set.

References

Collins, A., Musser, W. N., & Mason, R. (1991). Prospect theory and risk preferences of oregon seed producers. *American Journal of Agricultural Economics,* 429–435

Fishburn, P. C., & Kochenberger, G. A. (1979). Two-piece von Neuman–Morgenstern utility functions. *Decision Sciences, 10*, 503–518

Friedman, D. (1989). The S-shaped value function as a constrained optimum. *The American Economic Review,* 1243–1248

Galanter, E., & Pliner, P. (1974). Cross-modality matching of money against other continua. In H. R. Moskowitz et al. (Eds.), *In sensation and measurement*. (pp. 65–76). Dordrecht, The Netherlands: Reidel

Kahneman, D., & Tversky, A. (1979). Prospect theory. An analysis of decision under risk. *Econometrica, 47*(2), 263–291

Robson, A. J. (2002). Evolution and human nature. *Journal of Economic Perspectives, 16*(2), 89–106

Rojas, R. (1996). *Neural networks. A systematic introduction*. Berlin: Springer

Rust, R. T. (2004). *An evolutionary basis for prospect theory* (Working Paper). The Netherlands: Tilburg University Seminar Series

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323

Part III
Experimental Individual Decision Making

Individual Choice from a Convex Lottery Set: Experimental Evidence

T. Neugebauer

Keywords: Individual choice under risk · First order stochastic dominance · Modern portfolio theory · Prospect theory

1 Introduction

This paper is concerned with a simple pen-and-paper experiment on individual choice under risk, in which subjects choose a lottery from a convex set. In the reported experiment subjects face a choice of two risky lotteries and a degenerated one, and any linear combination of the three lotteries.[1] The distinguishing features of the design are as follows: The two risky lotteries perfectly negatively correlate with each other, implying the existence of a riskless combination of these lotteries. Furthermore, as this riskless combination of the risky lotteries yields a greater payoff than the degenerated lottery, all lotteries in the interior of the convex set are strictly dominated. Finally, the efficient frontier of the convex set includes lotteries that involve a possible loss.

The experimental design including these features is useful for the following reasons. First, we can test whether subjects make a rational choice or whether they choose dominated lotteries in this setting. In theories on rational choice, dominance is a normatively essential requirement (Levy, 1998; Luce & Marley, 2005; Starmer, 2000 survey the literature).[2] Violations of dominance have been observed in binary choice experiments where dominance was not transparent (Tversky & Kahneman, 1986),[3] and in portfolio selection experiments.[4] However, in contrast to these studies the present one involves a riskless alternative. Since sure events are evaluated differently from risky events, a different choice behavior than in the reported studies can be conceivable. Second, a couple of experiments on portfolio

T. Neugebauer
University of Luxembourg, Luxembourg School of Finance,
4 rue Albert Borschette, 1246 Luxembourg, Luxembourg
e-mail: Tibor.Neugebauer@uni.lu

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*.
Theory and Decision Library C 42.
© Springer-Verlag Berlin Heidelberg 2008

selection choice show that participants do not respond to correlations between alternatives when they make their investment decisions (Clemen & Reilly, 1999; Kroll & Levy, 1992; Kroll et al., 1988a, b; Lipe, 1998; Oehler, 1995; Weber & Camerer, 1998). The available empirical evidence for portfolio selection is also compatible with the view that people ignore covariance risk, as negligence of covariance risk leads to under-diversification in financial markets.[5] However, no study involves perfect negative correlation which, again, enables the decision maker to eliminate all risk (Elton & Gruber, 1981). Finally, most experimental studies involve only gains, but the literature has provided convincing evidence that people treat possible losses differently from gains. Hence, the choices in our experiment may diverge from the existing evidence for several reasons.

Below I describe three classroom experiments in which highly motivated subjects, who possess complete information about risk and return involved in the decision task, are asked individually to choose a linear combination of one riskless asset and two risky assets. The two risky assets are perfectly negatively correlated, thus allow the construction of a riskless portfolio with a greater return than the riskless asset. Under these simplified conditions, I examine whether the actual choices from the convex set approach efficiency, and if theoretical or psychological principles underlie the observed behavior.

In the three experiments, each subject was asked to allocate a 100% share between three lotteries and provide a free-form rationale for the decision. The first (Original) and second (High Stakes) experiments are identical but the latter affects a considerably more highly paid group of subjects. These experiments involve one decision only; no repeated choice was considered (see the discussion on maximization of the geometric mean in Kroll et al. 1988a). The third (Transparency) experiment features repeated choice: the first stage is identical to the first experiment, but at a second stage subjects are exposed to a more transparent presentation of the task. The three experiments were designed to investigate four major issues: (1) the effects of the perfect negative correlation between the risky alternatives, and of the existence of the riskless but dominated lottery; (2) the possibility to choose from a convex set; (3) the effects of higher stakes; and (4) the effects of transparent presentation of dominance.

Section 2 describes the tasks employed in both experiments. The opportunity set and the theoretical optimal decisions are presented in Sect. 3. Section 4 describes and discusses the results of the Original and the High Stakes experiments; Sect. 5 does the same for the Transparency Experiment. Concluding remarks are stated briefly in Sect. 6.

2 Experimental Design

2.1 Instructions

Subjects had to write their choice on a decision sheet which included the following instructions: "There are three assets A, B, C; assets A and B are risky, C is riskless. The payoffs generated by these assets depend on two equiprobable states, X and Y.

Table 1 Payoffs in the experiment

Your division of your endowment	Asset	State X	State Y
a:	A	3	-1
b:	B	-3	6
1-a-b:	C	1	1
	Payoff	$(3a - 3b + 1 - a - b)$	$(-a + 6b + 1 - a - b)$

Your task is to make a decision about the allocation of your investment capital towards each of these assets. Please record the shares that you allocate to these assets in the first column of the following table. The sum of the shares must yield 100%, and each share must be non-negative.

After you have made your choice, you are asked to toss a coin to determine the state to occur. Before you toss, you decide whether heads represents state X or state Y. Given you choose X, state X occurs if the face on the upside of the coin shows heads; if the coin shows tails, state Y occurs. Given you choose Y, heads and tails imply state Y and X, respectively.

The resulting payoffs are recorded in the table. After you have allocated the shares that composed your portfolio, please compute your payoff for both states X and Y and record these numbers in the last row of the table. The corresponding amount will be paid to you in cash after you have tossed the coin. Note: if the outcome of the gamble is negative you will also have to pay your dues to the experimenter. In fact, you do not have to take any risk. Asset C is riskless and guarantees a sure payoff of 1 token."

All students first made their choice, and then they were asked to briefly state the rationale for their decision in words on the record sheet. In the last line of their sheet corresponding to the one in Table 1, subjects wrote the payoffs they were about to receive in the states X and Y. If the payoff in one state was negative, subjects were prompted to put the exact amount of money on the table to show that they were willing to execute their decision before the toss of the coin. Three students chose an allocation that involved a negative payoff in state X or Y. One student actually received a negative payoff and had to pay the corresponding amount to the experimenter. When all had finished their statement, each student tossed a coin to determine his/her own payoff according to the allocation of shares in the Table 1, and received immediate cash payment. The task took about three quarters of an hour to complete.

3 Theoretical Considerations

Before we discuss the experimental results we present the theoretical predictions. Due to perfect negative correlation, a riskless combination of A and B exists. We anticipate that most subjects choose a positive share of the degenerated lottery C.

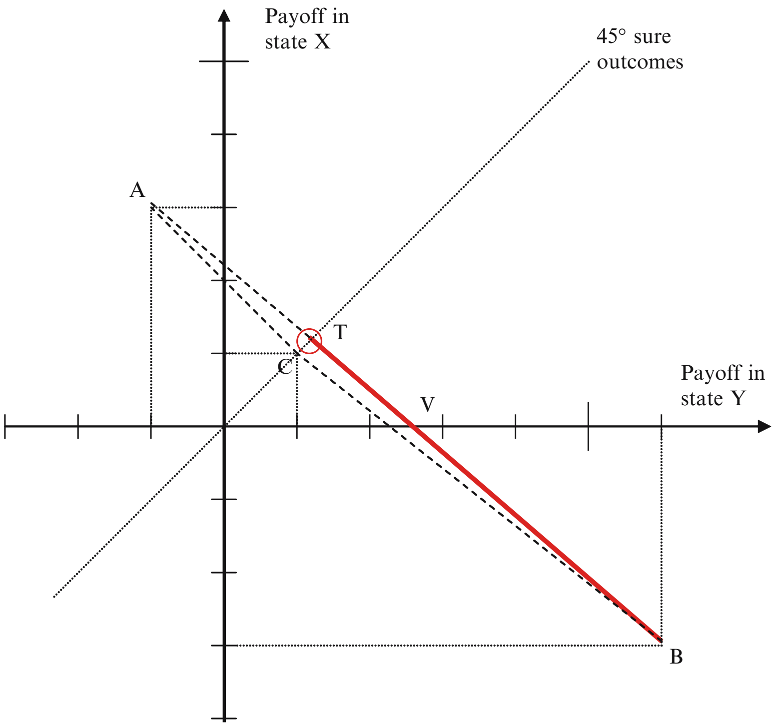


Fig. 1 Portfolio possibility set in the experiment

This decision is dominated, because the riskless portfolio constructed from the risky lotteries A and B at a ratio 9:4 induces a riskless return of $15/13 > 1$.[6] All allocations which involve a share greater than a = 9/13 towards the lottery A are dominated. The expected payoff of the three lotteries are $\{\mu_A; \mu_B; \mu_C\} = \{1; 1.5; 1\}$, the standard deviations of A and B are $\{\sigma_A; \sigma_B\} = \{2; 4.5\}$, the covariance ($\sigma_{AB} = -9$) and the correlation coefficient ($\rho_{AB} = -1$) between the risky lotteries. In Fig. 1, the plane included in the connecting lines of the coordinates A, B and C presents the convex set of lotteries one chooses from. The diagram represents the payoff space; the vertical axis measures the payoff in state X and the horizontal axis measures the payoff in state Y. The 45° line represents equal payoffs in both states, X and Y.

3.1 Testable Hypothesis: Efficient Frontier

The solid line in Fig. 1 represents all efficient choices. It connects lottery B with the riskless combination of A and B, corresponding to the coordinate (15/13; 15/13) denoted by T. The efficient choices involve all payoff maximizing combinations for a given amount of risk. Thus all indifference curves that maximize utility are tangent to this line. Note, the linear combinations that lie on the broken line which connects lottery A and T are dominated by the lotteries on the efficient line. Also, all choices in the interior of the set are dominated as they involve the strictly dominated lottery C. The choice of any combination on the efficient frontier is proposed in general by expected utility theory and in particular by mean-variance theory.[7]

3.2 Testable Hypothesis: Lossless Combination of Risky Lotteries

In prospect theory losses loom larger than gains. Hence, prospect theory as well as cumulative prospect theory would predict a choice on the efficient line between T and V for reasonable chosen loss aversion parameters. V represents the lottery involving the payoffs (2.5; 0) in states Y and X, i.e., V maximizes expected value in the domain of gains. In addition to these theories, other theories that weigh losses more than gains (for instance aspiration level theory with a positive aspiration level) would predict the positive segment of the efficient line.

3.3 Testable Hypothesis: Riskless Combination of Risky Lotteries

The riskless combination of risky lotteries, T, which generates a sure payoff of 15/13 is proposed by at least three theories: First, cumulative prospect theory involving the parameters estimated by Tversky and Kahneman (1992) would suggests this choice. Second, under the assumption of a perfect capital market, T would be the tangential portfolio suggested by the separation theorem and the CAPM. Third, safety first theory would also suggest this outcome, given that the safety first level is below 15/13.

4 Original and High Stake Experiment

4.1 Original Experiment

Fifteen students of the Behavioral Finance lecture in the summer term 2005 at the University of Hannover, Germany were asked to make their decision on a record sheet as displayed in Table 1. Subjects were no volunteers in the experiment, but completed the task as part of the lecture. One can assume that the students understood the task, since they all correctly calculated the payoffs for states X and Y on their record sheets. Most of them had seen the mean variance model in earlier courses. Despite the arguably small payoffs (one token equaled 1 €), students seemed highly motivated. Subjects were asked to imagine actually facing an honorable bet between millionaires.

Table 2 Individual choices (a, b and c in percentage) and stated rationale in the Original Experiment

ID	R(Y)	R(X)	a	b	c	Stated rationale explaining choice
1	0.90	1.20	30	10	60	A is less risky than B, I prefer the riskless asset
2	1.00	1.00	0	0	100	Assets A and B are very risky, I do not want to take any risk
3	1.05	1.10	35	15	50	B's variance is huge
4	1.33	1.00	67	33	0	perfect negative correlation, I try to estimate the optimal portfolio
5	1.75	0.50	25	25	50	Diversification
6	1.75	0.50	25	25	50	Payoffs are always positive
7	1.90	0.40	30	30	40	
8	2.00	0.33	33	33	33	
9	2.10	0.20	20	30	50	I always have a positive payoff
10	2.10	0.20	20	30	50	
11	2.20	0.20	40	40	20	My payoff is always positive
12	2.50	0	50	50	0	No loss, but relatively high gain possible
13	2.50	0	50	50	0	No loss, but relatively high gain possible
14	3.00	−0.50	25	50	25	Diversification with a tendency towards the more risky asset
15	6.00	−3.00	0	100	0	Take the risk, since little at stake. In case of higher stakes would decide differently.

Individual Choices in the Original Experiment

The individual choices of the original experiment are recorded in Table 2; percentages are rounded to the next integer and tokens are rounded to the next hundredth. The first column assigns an identification number to subjects; the second and third columns present the resulting returns in state X and Y, respectively. The third, fourth and fifth columns correspond to the allocation of shares to the lotteries A, B and C; reported numbers refer to percentages. For instance, individual #10 allocated a 30% share to A, 10% to B and 60% to C; the corresponding payoffs for the states X and Y were 1.20 and 0.90 tokens.

The expected payoffs among all subjects average at 1.17 tokens. This amount is only 0.02 tokens more than the payoff of the riskless combination T. The Wilcoxon (signed ranks) test indicates that the difference in payoff to the riskless choice is insignificant ($p > 0.5$). Only four subjects chose a lottery on the efficient frontier (#4, #12, #13, #15). Due to the students' choices of asset C, the average loss per subject is 0.06 tokens in expected terms; the 95% confidence interval extends from 0.03 to 0.08 tokens. The difference is significantly different from zero ($p < 0.002$, Wilcoxon test). The efficiency loss can be expressed in excess risk incurred by the students. On the efficient frontier, the risk to be incurred at an expected payoff of 1.17 tokens would be 0.26 tokens. In other words, the standard deviation which is 0.99 tokens is 0.73 tokens too large.

In fact, it would be good to know whether subjects have any reason for such non-rational choices. The last column of Table 2 records the stated rationale of the subject eventually abbreviated and translated to English. Subjects #7, #8, and #10 did not provide any statement. In fact, no salient rewards were connected to these statements. Subject #4 is the only one who states that perfectly negative correlation is involved, thus he presumably recognized the possibility of a riskless combination of assets A and B. All other subjects make no analogous statement, but rather refer to possible losses or to the wish to diversify. At least the statements of five subjects (#6, #9, #11, #12, #13) suggest that they do not want to incur any loss.

4.2 High Stake Experiment

Eighteen students of the course "Decision Making and Portfolio Choice" in the winter term 2006 at the University Hannover were asked to make their decision in the described experiment with tenfold payoff; i.e., each token equals 10 €. Kroll et al. (1988a) suggested that subjects may take the task more seriously and make different decisions if payoffs are multiplied by ten. The experiment was run at the end of the final lecture, 3 weeks ahead of students' final examination; expected utility and the mean-variance model were essential content in this course. All subjects were volunteers and no subject had previously participated in a comparable experiment.

Individual Choices in the High Stake Experiment

Corresponding to Tables 2 and 3 records subjects' choices, rationales and payoffs involved in the individual choices; the ID numbers are {#16, #17, ..., #33}, thereby taking into account the assigned IDs of the Original Experiment. The expected payoffs among subjects in the High Stake Experiment average 1.16 tokens.[8] This amount is only 0.01 tokens more than the payoff of the riskless combination, T, and 0.01 tokens less than in the Original Experiment. According to the Wilcoxon test, the difference in expected payoff is insignificant from the ones in T. Furthermore, the differences in choices and the differences in payoffs as measured in tokens are insignificant between the Original Experiment and the High Stake Experiment as the Mann Whitney test suggests.[9] In the High Stake experiment, five subjects choose a lottery on the efficient frontier. All these subjects {#29, #30, #31, #32, #33} choose the lossless corner lottery V. In summary, the tenfold increase in payoffs has no statistically notable effect on subjects' decisions. However, 13 subjects (72%) state that they do not want to earn a negative payoff, {#17, #20, #23, #24, .., #33}, and no subject chooses a lottery that involves a potentially negative payoff.

Table 3 Individual choices (a, b and c in percentage) and stated rationale in the High Stake Experiment

ID	r(Y)	r(X)	a	b	c	Stated rationale explaining choice
16	0.60	1.60	70	20	10	$r(Y) > r(X)$
17	0.70	1.40	40	10	50	No deposit payment. I am very risk averse
18	1.00	1.00	0	0	100	Don't invest in A, since $\mu = 10 = r(C)$. For B $\mu > 10$, but σ is too high
19	1.00	1.20	50	20	30	Risk relatively neutralized. In any case, I receive more than C
20	1.40	0.80	30	20	50	I am risk averse. I do not want to lose any, but I want a chance to earn some money
21	1.75	0.50	25	25	50	It is difficult to decide between A and B. I am risk neutral
22	1.80	0.40	10	20	70	
23	1.90	0.40	30	30	40	No Loss. Would C be greater, I would invest more in C
24	1.90	0.40	30	30	40	Because: C is riskless. B, A involve a possibility of loss or gain
25	2.00	0.33	33	33	33	No loss possible, asset A insures loss of B in X
26	2.10	0.20	20	30	50	No negative payoff, because budget = 0.50% chance of having a "high" payoff
27	2.40	0	30	40	30	All or nothing, but incur no losses
28	2.40	0	30	40	30	No risk of loss, but prospect of positive gain!
29	2.50	0	50	50	0	I do not feel like losing money
30	2.50	0	50	50	0	50% chance to receive something
31	2.50	0	50	50	0	Since one-shot gamble, I do not want to make a loss. If repeated I'd always choose B
32	2.50	0	50	50	0	In all states there is no loss
33	2.50	0	50	50	0	All or nothing

4.3 What Lottery Do Subjects Select from the Convex Set?

Since there are no significant differences between the data of the Original Experiment and the High Stake Experiment, the data are pooled in this section. The resulting 33 independent choices are displayed in Fig. 2 with respect to the shares *a* and *b* allocated to the risky lotteries A and B. In the figure, a small circle represents one observation, a double circle represents two choices, a triple circle three choices etc. Hence, seven subjects chose the 50:50 division of endowment toward the two risky lotteries inducing the payoffs 2.50 tokens or 0. As can be seen in the figure, most choices are on the 45° line or close to it; 14 choices involve a = b; 8 involve a > b; and 7 choices involve a < b; on average a = 1.03 b. The distribution around the 45° line is approximately symmetric.[10] In other words, the allocation to the risky lotteries is divided in equal shares among A and B. Although 34% of shares were assigned to C a glance at Fig. 2 makes it evident that it is not apt to say that the endowment is divided equally between all three lotteries. The general choice pattern suggested by the figure seems rather in line with the following heuristic: choose a share of the riskless lottery and allocate the remainder to the lottery that maximizes expected value among the lossless lotteries. In fact, this choice pattern is not rational

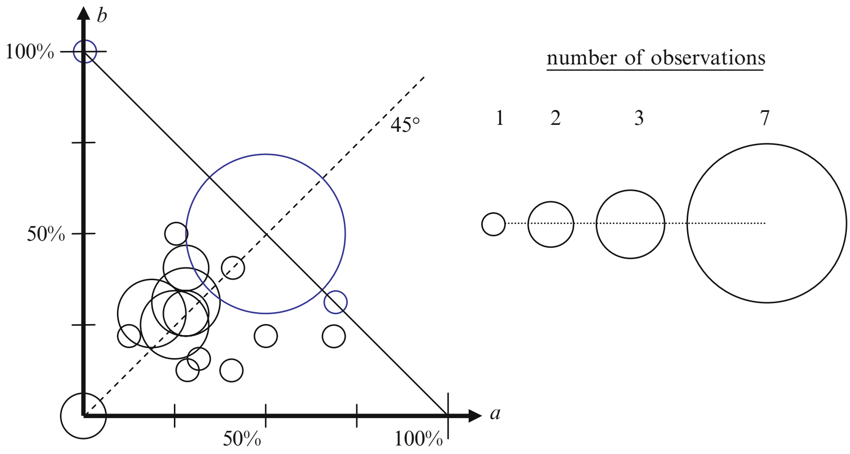


Fig. 2 Risky lottery share in the experiment

as it involves a dominated lottery, but the rationale agrees with aspiration level theory (Lopes, 1987) which forms the basis of behavioral portfolio theory (Shefrin & Statman, 2000).[11]

5 Transparency Experiment

5.1 The Experiment

Thirteen students of the Behavioral Finance lecture in the summer term 2006 at the University of Hannover voluntarily participated in the experiment. The experiment was run at the end of a lecture. Only some of the 64 students who attended the course had previous knowledge on portfolio choice and individual decision making. It is possible that some of the subjects had participated in one of the other above reported experiments. The participating subjects, in Table 4 identified by {#34, #35, ..., #46}, were first asked to make a choice according to Table 1; in accordance with the Original Experiment, one token equaled 1 € in the Transparency Experiment. When all subjects had made the first choice, they were asked to make a second choice on another record sheet. The second sheet presented 14 lotteries including lotteries C and V.[12] The other twelve lotteries corresponded to coordinates on the line that connects A and B in Fig. 1, dividing the line in 13 equal sized segments. Lotteries A and B were not included in the sheet and the states X and Y were interchanged to mask the decision problem; the two tasks in the experiment should not be easily identified as identical.[13] Hence, in the second decision task, the two riskless lotteries C and T were both exposed to the participants, such that dominance was transparent. Either the choice in the first task or the choice in the second task was played out for real; subjects tossed a coin to determine the relevant task.

Table 4 Individual choices in the Transparency Experiment

ID	r(X)	r(Y)		r(Y)	r(X)
34	1.60	0.40	→	1.15	1.15
35	1.00	1.00	→	1.15	1.15
36	1.00	1.00	→	1.15	1.15
37	1.00	1.00	→	1.15	1.15
38	1.00	1.00	→	1.88	0.58
39	0.60	1.80	→	2.23	0.23
40	0.30	2.00	→	2.23	0.23
41	−0.50	3.50	→	2.50	0
42	0	2.50	→	2.50	0
43	0	2.50	→	2.50	0
44	0	2.50	→	2.50	0
45	0	2.50	→	2.50	0
46	0	2.50	→	2.50	0

Individual Choices in the Transparency Experiment

Since independence of the data from the above experiments is not warranted and subjects did hardly state any rationale on their sheet at the end of the second task, this section focuses on the differences in choice between the first and the second task in the Transparency Experiment. Table 4 records on the left side the lotteries chosen in the first task, and the corresponding choices of the second task on the right. While in the first task the majority of decisions involved a dominated lottery, 0% of subjects chose a dominated lottery in the second task. This result confirms earlier experimental results of the non-dominated lottery choice where dominance was transparent (Birnbaum, 1998a; Tversky & Kahneman 1986).

It might be conceivable that subjects who have chosen a risky lottery in the first task may choose T when this lottery is transparently presented in the second task, because it is the efficient riskless payoff. However, the table of the individual choices in the Transparency Experiment reveals that subjects who chose lottery V in the first task did repeat their choice in the second task. Between tasks, subjects chose the same amount of risk; in the first task the average standard deviation between payoffs in state X and Y was 1.12 tokens, in the second task it was 1.10 tokens ($p > 0.5$, Mann Whitney test).

6 Summary

This paper has presented a simple individual choice experiment in a classroom setting, where subjects choose a portfolio from a convex set involving dominated lotteries and perfect negative correlation of payoffs. Most subjects had heard about expected utility theory and mean variance theory before, but the presentation of the task was different than the one from their textbooks. The correlation involved in the task allowed the construction of a riskless lottery, but hardly any subject identified the correlation. Hence, the present study shows that even under perfect negative correlation subjects neglect correlation. Moreover, most subjects choose a dominated lottery when dominance is not transparent. The fact that all subjects choose an efficient lottery when dominance is transparent seems to suggest that errors cause efficiency losses.[14]

The data on revealed preferences and the stated rationales of subjects suggest that subjects in the experiment choose a share risklessly and allocate the remainder of their endowment to the most risky lottery that involves only gains. The data thus support earlier experimental findings according to which subjects are loss averse rather than variance averse (Duxbury & Summers, 2004; Levy & Levy, 2001). This pattern seems to suggest that subjects want both, secure a non-negative payoff and achieve a high expected return. Though choices are only boundedly rational since resources are allocated to a dominated lottery, the behavior is in the spirit of aspiration level theory (Lopes, 1987).

Acknowledgement I acknowledge helpful comments by Klaus Spremann and seminar participants at the FUR conference Rome 2006, in particular Guido Baltussen, Michael Birnbaum, Jerome Busemeyer, Thierry Post, Uli Schmidt, and Stefan Traub.

Notes

[1]The available combinations represent mixtures of consequences at a fixed probability of one half. This approach contrasts to the one presented in Sopher & Narramore (2000) where subjects could mix probabilities over fixed consequences.

[2]Systematic dominance violations refute a large class of descriptive models including rank-dependent utility theory (Diecidue & Wakker, 2001; Quiggin, 1985, 1993), rank and sign dependent utility theory (Luce & Fishburn, 1991, 1995), cumulative prospect theory (Gonzalez & Wu, 1999; Starmer & Sugden, 1989; Tversky & Kahneman, 1992; Tversky & Wakker, 1995; Wakker & Tversky, 1993; Wu & Gonzalez, 1996), lottery dependent utility theory (Becker & Sarin, 1987), aspiration level theory (Lopes, 1987; Lopes & Oden, 1999), mean-variance theory (Lintner, 1965; Markowitz, 1952; Mossin, 1966; Sharpe, 1964; Tobin, 1968), safety first theory (Kataoka, 1963; Roy, 1952; Telser, 1955), and generalized utility theory (Machina, 1982).

[3]Birnbaum and associates present recent evidence for systematic violations of first order stochastic dominance (Birnbaum, 1997, 1999a, b, 2004a, b; Birnbaum & Martin, 2003; Birnbaum & Navarette, 1998; Birnbaum et al. 1999). Diederich & Busemeyer (1999) report also violation of first order stochastic dominance in a repeated setting where dominance is not transparent.

[4]Kroll & Levy (1992) and Kroll et al. (1988a, b) find violations of second order stochastic dominance, and Baltussen and Post (2005) report violations of first order stochastic dominance.

[5]Empirical studies have documented that private investors under-diversify and frequently invest only in one to two securities or they split their wealth equally over all available assets or funds (Bernatzi, 2001; Bertaut, 1998; Blume & Friend, 1975; Blume et al., 1974; Cohn, Lewellen, & Lease, 1975; Guiso, Japelli, & Terlizze, 1996; Heaton & Lucas, 2000; Joos & Kilka, 1999; Kelly, 1994; Perraudin & Sorensen, 2000; Samuelson & Zeckhauser, 1988; Statman, 1987). The latter investment approach is known as the 1/n-heuristic or naïve diversification (Benartzi & Thaler, 2001), and even financial advisors support corresponding investment decisions (Canner, Mankiw, & Weil 1997; Elton & Gruber, 2000; Fisher & Statman, 1997a, b; Siebenmorgen, Weber, & Weber 2001). Siebenmorgen and Weber provide a remarkable fit between a behavioral model that assumes the negligence of covariance risk in the objective function and their own data as well as the one of Canner et al.

[6]To compute the riskless combination of lotteries A and B, equalize the payoffs in states X and Y: $3a - 3b + 1 - a - b = -a + 6b + 1 - a - b \Leftrightarrow a/b = 9/4$, where *a* and *b* denote the shares allocated to A and B.

[7]Other theories predict also choices on the efficient frontier (general form of cumulative prospect theory, prospect theory, rank-dependent utility theory, rank and sign dependent utility theory, lottery dependent utility theory, aspiration level theory and generalized utility theory).

[8]The average loss in payoff is 0.06 tokens, i.e., the same as in the Original Experiment. The difference is significantly different from zero ($p = 0.001$, Wilcoxon test). Efficiency is 95% in both experiments. The standard deviation for the expected payoff of 1.16 tokens on the efficient line would be 0.05 tokens, 0.76 tokens less than the observed standard deviation. The observed standard deviations of the chosen lotteries are statistically undistinguishable between Original and High Stake Experiment ($p = 0.957$, Mann Whitney test).

[9]With the Mann Whitney test I compared the allocations of shares to A, B and C (i.e., the shares a, b and c), the payoffs in states X, Y and the expected payoffs, both in tokens. Furthermore, I checked on the equality of the variances between the two experiments by means of the non-parametric test of Talwar & Gentle (1977). All probability values are greater than $p > 0.8$.

[10]The Wilcoxon signed ranks test for symmetry returns a probability value of $p = 0.513$.

[11]Aspiration level theory supports also both, the purchase of insurance and the purchase of lottery tickets.

[12]Although the task involved the choice of a linear combination from 14 lotteries, actually no subject chose a combination of more than two lotteries.

[13]The procedure tried to avoid that students made the same choice in both tasks just on grounds of consistency. Some subjects were debriefed after the experiment; these subjects had not noticed that the tasks were basically identical not regarding the risky lotteries A and B.

[14]A couple of models study the role of errors in individual decision making (Camerer & Ho, 1994; Harless & Camerer, 1994; Hey, 1995; Hey & Orme, 1994; Loomes & Sugden, 1995; Schmidt & Neugebauer, 2007).

References

Baltussen, G., & Post, G. T. (2005). Irrational Diversification; An experimental investigation of portfolio construction. (Mimeo, Erasmus University Rotterdam)

Becker, J. L., & Sarin, R. K. (1987). Lottery dependent utility. *Management Science, 33*, 1367–1382

Bernatzi, S. (2001). Excessive extrapolation and the allocation of 401(k) accounts to company stock. *The Journal of Finance, 56*, 1747–1764

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in retirement saving plans. *The American Economic Review, 91*, 79–98

Bertaut, C. C. (1998). Stockholding behavior of US households: Evidence from 1983–1989 survey of consumer finances. *Review of Economics and Statistics, 80*, 263–275

Birnbaum, M. H. (1997). Violations of monotonicity in judgment and decision making. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 73–100). Mahwah, NJ: Erlbaum

Birnbaum, M. H. (1999a). Paradoxes of Allais, stochastic dominance, and decision weights. In J. Shanteau, B. Mellers, & D. Schum (Eds.), *Decision science and technology: Reflections on the contributions of Ward Edwards* (pp. 27–52). Boston, MA: Kluwer

Birnbaum, M. H. (1999b). Testing critical properties of decision making on the Internet. *Psychological Science, 10*, 399–407

Birnbaum, M. H. (2004a). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48*, 87–106

Birnbaum, M. H. (2004b). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes, 95*, 40–65

Birnbaum, M. H., & Martin, T. (2003). Generalization across people, procedures, and predictions: Violations of stochastic dominance and coalescing. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on decision research* (pp. 84–107). New York, NY: Cambridge University Press

Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty, 17*, 49–78

Birnbaum, M. H., Patton, J. N., & Lott, M. K. (1999). Evidence against rank-dependent utility theories: Violations of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes, 77*, 44–83

Blume, M., Crockett, J., & Friend, I. (1974). Stock-Ownership in the United States: Characteristics and Trends. *Survey of Current Business*, 54(11), 16–40

Blume, M. E., & Friend, I. (1975). The asset structure of individual portfolios and some implications for utility functions. *Journal of Finance, 30*, 585–603

Camerer, C. F., & Ho, H. T. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty, 8*, 167–196

Canner, N., Mankiw, N. G., & Weil, D. N. (1997). An asset allocation puzzle. *The American Economic Review, 87*, 181–191

Clemen, R. T., & Reilly, T. (1999). Correlations and copulas for decision and risk analysis. *Management Science, 45*, 208–224

Cohn, R. A., Lewellen, W. G., Lease, R. C., & Schlarbaum, G. G. (1975). Individual investor risk aversion and investment portfolio composition. *Journal of Finance, 30*, 605–620

Diecidue, E., & Wakker, P. P. (2001). On the intuition of rank-dependent utility. *Journal of Risk and Uncertainty, 23*, 281–298

Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic dominance principle of decision making. *Psychological Science, 10*, 353–359

Duxbury, D., & Summers, B. (2004). Financial risk perception – Are individuals variance averse or loss averse? *Economics Letters, 84*, 21–28

Elton, E. J., & Gruber, M. J. (1981). *Modern portfolio theory and investment analysis*. New York, NY: Wiley

Elton, E. J., & Gruber, M. J. (2000). The rationality of asset allocation recommendations. *Journal of Financial and Quantitative Analysis, 35*, 27–41

Fisher, K. L., & Statman, M. (1997a). The mean-variance-optimization puzzle: Security portfolios and food portfolios. *Financial Analysts Journal, 53*, 41–50

Fisher, K. L., & Statman, M. (1997b). Investment advice from mutual fund companies – Closer to the Talmud than to Markowitz. *The Journal of Portfolio Management, 53*, 9–25

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*, 129–166

Guiso, L., Japelli, T., & Terlizze, D. (1996). Income risk, borrowing constraints and portfolio choice. *American Economic Review, 86*, 158–172

Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica, 62*, 1251–1289

Heaton, J., & Lucas, D. (2000). Portfolio choice and asset prices: The importance of entrepreneurial risk. *Journal of Finance, 55*, 1163–1198

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review, 39*, 633–640

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica, 62*, 1291–1326

Humphrey, S. J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty, 11*, 263–274

Joos, C. M., & Kilka, M. (1999). Sind Aktienportfolios privater Anleger ausreichend diversifiziert? Die Bank 12/99: 862–866
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291
Kataoka, S. (1963). A stochastic programming model. *Econometrica, 31*, 181–196
Kelly, M. (1994). All their eggs in one basket: Portfolio diversification of US households. *Journal of Economic Behavior and Organization, 27*, 87–96
Kroll, Y., & Levy, H. (1992). Further tests of the separation theorem and the capital asset pricing model. *The American Economic Review, 82*, 664–670
Kroll, Y., Levy, H., & Rapoport, A. (1988a). Experimental Tests of the Separation Theorem and the Capital Asset Pricing Model. *The American Economic Review, 78*, 500–519
Kroll, Y., Levy, H., & Rapoport, A. (1988b). Experimental Tests of the Mean-Variance Model for Portfolio Selection. *Organizational Behavior and Human Decision Processes, 42*, 388–410
Levy, H. (1998). *Stochastic dominance: Investment decision making under uncertainty*. Boston, MA: Kluwer
Levy, M., & Levy, H. (2001). Testing for risk aversion: a stochastic dominance approach. *Economics Letters, 71*, 233–240
Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics, 47*, 13–37
Lipe, M. G. (1998). Individual investors' risk judgments and investment decisions: The impact of accounting and market data. *Accounting, Organizations and Society, 23*, 625–640
Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review, 39*, 641–648
Lopes, L. L. (1987). Between hope and fear: The psychology of risk. In: Leonard Berkowitz, (ed.), *Advances in Experimental Social Psychology*. San Diego, CA: Academic Press, 255–295
Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology, 43*, 286–313
Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first order gambles. *Journal of Risk and Uncertainty, 4*, 29–59
Luce, R. D., & Fishburn, P. C. (1995). A note on deriving rank-dependent utility using additive joint receipts. *Journal of Risk and Uncertainty, 11*, 5–16
Luce, R. D., & Marley, A. A. J. (2005). Ranked additive utility representations of gambles: Old and new axiomatizations. *Journal of Risk and Uncertainty, 30*, 21–62
Machina, M. J. (1982). 'Expected utility' theory without the independence axiom. *Econometrica, 50*, 277–323
Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance, 7*, 77–91
Marley, A. A. J., & Luce, R. D. (2005). Independence properties vis à-vis several utility representations. *Theory and Decision, 58*(1), 77–143
Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica, 34*, 261–276
Oehler, A. (1995). *Die Erklärung des Verhaltens privater Anleger*. Stuttgart: Schaeffer-Poeschel
Perraudin, W. R. M., & Sorensen, B. E. (2000). The demand for risky assets: Sample selection and household portfolios. *Journal of Econometrics, 97*, 117–144
Quiggin, J. (1985). Subjective utility, anticipated utility, and the Allais paradox. *Organizational Behavior and Human Decision Processes, 35*, 94–101
Quiggin, J. (1993). *Generalized expected utility theory: The rank-dependent model*. Boston: Kluwer
Roy, A. D. (1952). Safety-first and the holding of assets. *Econometrica, 20*, 431–449
Samuelson, W., & Zeckhauser, R. J. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1*(1), 7–59
Schmidt, U., & Neugebauer, T. (2007). Testing expected utility in the presence of errors. *Economic Journal, 117*, 470–485
Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*(3), 425–442

Shefrin, H., & Statman, M. (2000). Behavioral Portfolio Theory. *Journal of Financial and Quantitative Analysis, 35*(2), 127–151

Siebenmorgen, N., Weber, E. U., & Weber, M. (2001). *A behavioral approach to the asset allocation puzzle*. (Mimeo, Universität Mannheim)

Sopher, B., & Narramore, J. M. (2000). Stochastic choice and consistency in decision making under risk: An experimental study. *Theory and Decision, 48*(4), 323–350

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 38*, 332–382

Starmer, C., & Sugden, R. (1989). Violations of the independence axiom in common ratio problems: An experimental test of some competing hypotheses. *Annals of Operations Research, 19*, 79–101

Statman, M. (1987). How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis, 22*, 353–363

Talwar, P. P., & Gentle, J. E. (1977). A robust test for the homogeneity of scales. *Communications in Statistics – Theory and Methods, A6*, 363–369

Telser, L. (1955). Safety-first and hedging. *Review of Economic Studies, 23*, 1–16

Tobin, T. (1968). Liquidity preferences as a behavior toward risk. *Review of Economic Studies, 25*, 65–85

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59*(4), 251–278

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323

Tversky, A., & Wakker, P. (1995). Risk attitudes and decision weights. *Econometrica, 63*, 1255–1280

Wakker, P., & Tversky, A. (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty, 7*, 147–176

Weber, M., & Camerer, C. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior and Organization, 33*, 167–184

Wu, G. (1994). An empirical test of ordinal independence. *Journal of Risk and Uncertainty, 9*, 39–60

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science, 42*, 1676–1690

Temptations and Dynamic Consistency

E. Carbone

Keywords: Hyperbolic discounting · Experiments · Spot markets · Forward markets

1 Introduction

The objective of this paper is to test an implication of the quasi-hyperbolic model of discounting by implementing an experiment on temptations. This implication is that people choose more investment goods than temptation goods when they plan, but choose more temptation goods than investment goods when they consume on the spot. This effect, which has been called the *immediacy effect* by Read et al. (1999), occurs when an individual behaves according to the quasi-hyperbolic discounting model, but not when an individual behaves according to the exponential discounting model.

The immediacy effect has been tested by Read et al. (1999) and by Read & Van Leeuwen (1998) in a particular experimental setting. These authors conclude from their experiments that the immediacy effect exists. However, the results of the experiment reported in this current paper – in a different experimental setting – do not confirm this effect. The main point of this current paper is not only to show that the immediacy effect is not observed in our experiment, but also to compare our results to those of Read et al., and try to understand why they are different. We will argue that these differences are due to important and relevant differences in the experimental design.

The paper is organised as follows. In the next section we briefly outline the exponential and quasi-hyperbolic models of discounting. We then apply these two

E. Carbone
Università di Bari, Facoltà di Economia, Via Camillo Rosalba 53
e-mail: e.carbone@dse.uniba.it

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.

© Springer-Verlag Berlin Heidelberg 2008

theories in a particular context and show that they may lead to different predictions. Section 3 we describe our experiment which was designed to try and distinguish whether individuals behave according to the exponential model or according to the quasi-hyperbolic model. Section 4 contains an analysis of the data analysis and the final section concludes.

2 The Theory Tested

The mainstream model of intertemporal choice in the economics literature is the Discounted Utility Model – in which the utilities of future consumption are discounted to the present. The Discounted Utility Model is typically implemented with an exponential discount function. According to this *exponential discounting model*, an individual discounts the future utility of consumption using a constant discount factor δ, so that the utility at time t of a stream of consumption $c_t, c_{t+1}, \ldots, c_T$ is given by the expression:

$$u(c_t) + \delta u(c_{t+1}) + \delta^2 u(c_{t+2}) + \cdots + \delta^{T-t+1} u(c_T) \tag{1}$$

Phelps & Pollak (1968) introduced the concept of quasi-hyperbolic discounting. This has recently been ‘re-discovered’ and extended by Laibson (1997). The quasi-hyperbolic discounting model is built on the idea (reinforced by empirical evidence) that consumers have a higher discount rate between the present and the following period than between any two adjacent subsequent periods. This *quasi-hyperbolic discounting model* implies that the utility at time t of a stream of consumption $c_t, c_{t+1}, \ldots, c_T$ is given by the expression:

$$u(c_t) + \beta[\delta u(c_{t+1}) + \delta^2 u(c_{t+2}) + \cdots + \delta^{T-t+1} u(c_T)] \tag{2}$$

The existence of the parameter β (if it is not equal to 1) distinguishes this model from the exponential model. Note that, as viewed from period t, the individual discounts the utility of period $t+s+1$ consumption relative to the utility of period $t+s$ consumption by δ, whereas, as viewed from period $t+s$, the individual discounts the utility of period $t+s+1$ consumption relative to the utility of period $t+s$ consumption by $\beta\delta$. Thus the *relative* discount rate varies according to the time of the comparison. This implies the possibility of time inconsistency.

The hypothesis we test is an implication of the hyperbolic model combined with the use of *temptation* goods or activities and *investment* goods or activities. Let us define an investment good or activity as a good which has current costs and future benefits, while a temptation good or activity is one that has current benefits and future costs. Read et al. (1999) use the terms *vices* and *virtues*: their virtues are our investment goods and activities; and their vices are our temptation goods and activities. Read et al. note (p. 259), “Someone who discounts the future in a hyperbolic or quasi-hyperbolic manner will be likely to prefer an immediate vice over an immediate virtue, since the vice offers a larger reward in the present. The same individual, however, might well take the virtue if both are delayed, since in this case the initial reward offered will no longer receive disproportionate weight". This is the hypothesis that we test in our experiment.

Let us state this more formally. Suppose that in the case of a temptation good the cost is paid a period after the benefit is received, while in the case of an investment good the benefit is received the period after the cost is paid. If an individual with hyperbolic discounting has to choose in the present, then, if he or she chooses the temptation (investment) good, the present benefit (cost) will not be discounted, while the future cost (benefit) will be discounted by $\beta\delta$. However, if the same individual has to decide *now* about what to consume in the following period, then, if he or she chooses the temptation (investment) good, then the present benefit (cost) will be discounted by $\beta\delta$, while the future cost (benefit) will be discounted by $\beta\delta^2$. The *relative* discounting changes according to the time that the decision is made. To illustrate this with a concrete example, we follow one given by Read et al. Assume that $\beta = 0.5$ and $\delta = 1$, and that the utility stream is (25, 200) for the virtue and (100, 100) for the vice, where the first entry represents the utility received when the vice or virtue is consumed and the second entry the utility in the period following. The individual should choose according to the discounted values of the two streams. If consumption follows immediately after the decision then the relevant discounted utilities are $125\,(= 25 + 0.5 \times 200)$ for the virtue and $150\,(= 100 + 0.5 \times 100)$ for the vice. The vice will be preferred. On the other hand, if the consumption can only take place one period after the decision, then the relevant discounted utilities are $112.5\,(= 0 + 0.5 \times (25 + 200))$ and $100\,(= 0 + 0.5 \times (100 + 100))$. Now the virtue will be preferred.

We now apply these ideas in a different context. Consider two different market situations: the first in which there is a spot market; and the second in which there is a forward market. In both markets good and activities are sold. In the spot market participants can consume at the same time as they buy. However, in the forward market, participants have to order the goods in advance of their consumption. The effect described above, called by Read et al. the *immediacy effect*, leads to the following prediction concerning differences in behaviour between the spot market and the forward market: people will chose more investment (temptation) goods or activities in the forward (spot) market than in the spot (forward) market. Of course, as Read et al. say, the possibility of observing this prediction depends on the relative desirability of the temptation and investment goods and activities.

3 The Experimental Design

Read et al.'s experimental setting required subjects to attend on two occasions. On the first occasion subjects were asked what they would like to consume (from a list of possibilities) when they returned on the second occasion. When they did actually return on the second occasion, they were asked again what they wanted to consume (from the same list) on that occasion. Note that the responses of the subjects on the first occasion were not implemented, so that subjects could well – and in fact did – change their minds. Indeed, that was the crucial finding of the experiment: de facto on the second occasion subjects consumed more vices than they had said (on the first occasion) that they wanted to consume on the second occasion. Read et al. take this as confirmation of the *immediacy effect*. For reasons that we discuss later, we prefer a different experimental setting. This is described below and takes its cue from the application of the quasi-hyperbolic model to the case of spot and forward markets discussed at the end of the section above.

The experiment was conducted at Economia Sperimentale al Sud d'Europa (ESSE) Lab at the University of Bari in Italy. The experiment consisted of eight sessions; four sessions were run at the end of January 2005 and the other four at the end of May 2005. The experiment was advertised through a leaflet distributed by hand in the Faculty of Economics at the University of Bari or sent by e-mail to a list of people that had participated in previous experiments. This leaflet (available on request) informed people that the experiment would last 5 h, that the participants could bring with them their textbooks, that during the experiment they could read magazines, play videogames and so on, and at the end of the experiment they would receive 50 € less what they had spent during the experiment.

There were two separate treatments in the experiment: a spot market treatment and a forward market treatment. Four of the sessions used the spot market treatment and four the forward market treatment. In each treatment a range of investment goods and activities and a range of temptation goods and activities were available for sale (the complete list of goods and activities is in the instruction reported in Appendix 1). In the spot market treatment the participants had 15 min to read the instructions (simultaneously read aloud by the experimenter) and subsequently 4 h and 45 min to buy and consume goods and activities that they wanted. In the forward market treatment the participants had 15 min to read the instructions (simultaneously read aloud by the experimenter) and order the goods and activities they wanted to consume during the subsequent 4 h and 45 min; crucially no other goods or activities could be bought after the first 15 min and the subjects were fully informed that that would be the case. During the ensuing 4 h and 45 min the subjects were brought the goods and activities that they previously had ordered.

In neither treatment could participants talk to each other during the experiment. Nor could they indulge in any other activity (other than doing nothing) other than those available in the experiment; nor could they consume any other good other than those available in the experiment. At the end of the experiment the participants were paid 50 € less what they spent during the experiment; All the goods were sold at half price of the faculty bar prices. All the activities were sold at 1 centesimo (of euro) per minute. Goods and activities bought during the experiment, but not consumed, could not be taken away or refunded at the end of the experiment.

At the end of the experiment participants answered a questionnaire, which allowed us to elicit some demographic information. Additionally in the questionnaire each subject had to rate each good in term of immediate pleasure and delayed benefit. We used these ratings to calculate temptation indices and investment indices for the various goods and activities (both individual – to the subject – and aggregate – to the participants as a whole). We shall say more about these indices shortly.

The consent form, the instructions, the debriefing questionnaire and the debriefing statement are reported in Appendix 1.

4 The Data Analysis

Recall that the prediction of our model is that people with hyperbolic discounting would consume more temptation goods in the spot market than in the forward market, and consume less investment goods in the spot market than in the forward market. We now proceed to a test of this prediction.

We describe the nature of our empirical investigation, after introducing some notation. In this, we refer to individual goods and activity with the subscript i and individual subjects with the subscript j. In our experiment there was a total of 24 goods and activities so i ranges from 1 to 24 and there was a total of 80 subjects so j ranges from 1 to 80.

The variables we observed during the experiment are the following:

e_{ij}: the expenditure on good and activity i by individual j. From these we can derive the total expenditure on good or activity i: $e_i = \sum_{j=1}^{80} e_{ij}$ and the total overall expenditure on all goods and activities of all individuals $e = \sum_{i=1}^{24} e_i$.

T_{ij}: the temptation felt by individual j with respect to good or activity i. This was measured by the response of the individual in the questionnaire to the question "Does this good or activity give you immediate pleasure? Is it fun, tasty or pleasurable". The variable T_{ij} was coded 0, 1 or 2 according as the response was "none", "a little" or "a lot".

I_{ij}: the "investability" felt by individual j with respect to good or activity i. This was measure by the response of the individual in the questionnaire to the question "Does this good or activity have benefits that last at least a few days? Is it healthy or educational?". The variable I_{ij} was coded 0, 1 or 2 according as the response was "none", "a little" or "a lot". This measures how much investment value there was to that individual of that good or activity.

T_i and I_i: *average* measures of temptation and investability of each good or activity defined as follows:

$$T_i = \frac{1}{80}\sum_{j=1}^{80} T_{ij} \text{ and } I_i = \frac{1}{80}\sum_{j=1}^{80} I_{ij}$$

F : a dummy for the forward market (forward $F = 1$ and spot $F = 0$).

In addition we gathered the following demographic data: whether the subject was a student; whether the subject was employed; the subject's annual food expenditure; his or her age; his or her height; his or her gender; his or her weight.

The first equation estimated is the following:

$$e_{ij} = \underset{(0.085)}{0.002} + \underset{(4.962)}{0.084T_{ij}} + \underset{(6.459)}{0.1171_{ij}} + \underset{(6.677)}{0.174 \quad F_j} - \underset{(-2.395)}{0.0053F_j \quad T_{ij}} - \underset{(-6.325)}{0.147F_jI_{ij}} \quad (i{=}1.24, j{=}1.80)$$

Note that the above equation implies:

$$\begin{array}{ll} \text{Spot market } (F = 0) & e = 0.002 + 0.0084T + 0.1171 \\ \text{Forward market } (F = 1) & e = 0.176 + 0.031T - 0.0301 \end{array}$$

When we substitute in the three possible values of the temptation index and the three possible values of the investment index we get the following table.

Spot market			Forward market		
T	I	e	*T*	I	e
0	0	0.002	0	0	0.176
1	0	0.086	1	0	0.207
2	0	0.170	2	0	0.238
0	0	0.002	0	0	0.176
0	1	0.119	0	1	0.146
0	2	0.236	0	2	0.116

This table tell us that when the temptation index is positive (1 or 2) and the investment index is zero that expenditure in the forward market is higher than the expenditure in the spot market. This goes against the predictions of the hyperbolic model. However, when the investment index is 1(2) and the temptation index is zero the expenditure in the spot market is lower (higher) than the expenditure in the forward market – so this evidence partly supports the quasi-hyperbolic prediction and partly refutes it.

An alternative, and perhaps more direct and robust, way of analysing the data is the following. Let us weight expenditures by their temptation and investment content. We have the temptation/investment coefficients from the questionnaire and depending on how we use them we can calculate weighted expenditure on temptation and investment goods and activities. Ideally we would like to say that expenditure on a particular good or activity is $x\%$ temptation and $(100 - x)\%$ investment – with the obvious extremes 100% temptation and 100% investment. The first of these occurs when the subject replies "a lot" to the temptation question and "none" to the investment question, and the second when the subject replies "none" to the temptation question and "a lot" to the investment question. Also it is clear when the subject replies "a little" to both questions that the good or activity is 50% temptation and 50% investment. However other cases are not so clear, and we have to decide a mapping from these answers on the questionnaires to the temptation and investment content of the expenditure. Consider the following mapping:

Temptation coefficients				Investment coefficients		
1/2	0	1	2	0	1	2
0	0.00	0.00	0.00	0.00	0.50	1.00
1	0.50	0.50	0.50	0.00	0.50	1.00
2	1.00	1.00	1.00	0.00	0.50	1.00

We note that, with this mapping, if the temptation index is 0, 1, 2 then the expenditure is 0, 50, 100% temptation, irrespective of the investment index. It the investment index is 0, 1, 2 then the expenditure is 0, 50, 100% investment, irrespective of the temptation index. However if both indexes are 2 this mapping implies that the expenditure is allocated 100% to temptation and 100% to investment – and hence is double counted. Clearly there is no ‘correct’ mapping, particularly if an individual regards a good or activity as providing both immediate and delayed pleasures. However, using this mapping (at an individual level using individual responses to the questionnaire to weight the individual expenditures) we get the following implied weighted expenditure percentages:

	Temptation	Investment
Spot	58	42
Forward	59	41

This shows that the expenditure on temptation goods and activities in the forward market is slightly higher than in the spot market, and that the expenditure on investment goods and activities is slightly higher in the spot then in the forward market. This is the opposite of that predicted by the quasi-hyperbolic model.

Of course, we may have chosen a bad mapping, but the above result appears robust. For the record, in Appendix 2, we include other mappings and their implied weighted expenditures. The implications do not seem particularly sensitive to the mapping.

5 Conclusions

The prediction of the quasi-hyperbolic model we tested in this experiment is not confirmed by the data. This prediction is that people choose more investment goods than temptation goods when they plan, but choose more temptation goods than investment goods when they consume on the spot.

To detect the plan that people make, and to compare that with what people consume on the spot, we used an experimental design in which there were two treatments: one containing a forward market and the other a spot market. We do not observe the expenditure reversal predicted by the hyperbolic model. What is observed, on the contrary, is that expenditure on temptation goods in the forward market is higher than in the spot market, and that expenditure on investment goods is lower in the forward market respect than in the spot market.

However, we should not dismiss immediately the quasi-hyperbolic discounting model on the basis of these results. There are two features of the experiment that may have caused this contrary result: one feature is the delay between the moment that people choose and the moment that people consume. One could argue that the probability that the preference reversal happens is inversely related to the temporal distance between the decision and the moment of consumption. In this experiment people would implement their plan soon after having chosen (they had a quarter of an hour to read the instruction and choose). We could, in order to put more time between the decision and the implementation of the plan, have implemented an experimental design in which the participants in the experiment had to go away and return on a second occasion. This would have created other problems – not least that the participants could have prepared for the second experimental occasion by eating and drinking in advance and coming with prepared reading.

The second problem in our experimental design is that which might be called the commitment or insurance effect: if people are risk averse and have to decide in advance what they are going to consume, they may choose more goods that they really want to avoid ending up without doing anything or eating less that they would desire. So they ordered more goods and activities they wanted – simply as insurance. It is not clear how one might control for this effect. However, our design has advantages over that of Read et al. In their design, there appear to be two problems, both related to the decision at the second stage. First, and de facto, subjects were allowed to change their minds on the second occasion – so that any plan that they formulated could be changed. Second, and as a consequence of this, if the subjects knew on the first occasion that they would be allowed to change their minds on the second occasion, it is not clear in what sense their stated plans at the first stage were in fact their true plans.

However, and this is still something to be explained, our experiment shows that we do not find any confirmation of the prediction of the quasi-hyperbolic model. We might have expected to find a modest effect, but, in fact, the movement of expenditure goes in the wrong direction.

Appendix 1

The Consent Form, the Instructions and the Debriefing Questionnaire and Statement

Consent Form

Professor Enrica Carbone of the Faculty of Economics at The University of Bari is conducting a study on how people make decisions in real-world environments. You may participate in this study only once.

You will receive 50 €, which you can spend on food or activities during the experiment. Whatever you don't spend you can take home with you at the end of the experiment. Your total participation should take around 5 h.

The data collected in this study will be used for economics research. Though nothing in the experiment is personally sensitive, we have taken steps to assure your anonymity. This anonymous data will be analyzed in future research.

I have read these instructions and agree to participate in the study:

________________________ (signature and date).

I Instructions for the Experiment

This experiment lasts 5 h. During the first 15 min we will explain the instructions and show you the goods and activities that can be purchased. During this time you will also be able to *select* the goods and activities that you want to purchase. After this, you will have 4 h and 45 min to *use* the goods and activities that you purchased. Goods and activities purchased but not consumed during the course of the experiment cannot be taken with you at the end of the experiment or refunded.

The *selection* and *purchase* of goods and activities can only be done from 12.30 to 12.45. To purchase a good or an activity, fill in the quantity on this page and the next page. After 12.45 no changes can be made. During the experiment, participants can speak only with the experimenter and her assistants. Please do not talk to other participants.

The prices for goods and activities are listed below. If you order something (like a cup of espressino), it will be delivered to you whenever you ask for it during the experiment. For example, if you purchase two cups of coffee, you have purchased the right to tell the experimenter to bring you a cup of coffee twice during the experiment. Remember that goods and activities purchased but not consumed during the course of the experiment cannot be taken with you at the end of the experiment or refunded.

We first list food items from the bar at the Faculty of Economics. Write the number you want (or "0").

- **Espressino coffee *24 centesimi*:** I agree to purchase ____ cups of espressino.
- **Espresso coffee *23 centesimi*:** I agree to purchase ____ cups of espresso.
- **Chocolate Snack *50 centesimi*:** I agree to purchase ____ chocolate snacks.
- **Focaccia *90 centesimi*:** I agree to purchase ____ focaccia.
- **Sandwich *90 centesimi*:** I agree to purchase ____ sandwiches.
- **Oransoda *40 centesimi*:** I agree to purchase ____ oransodas.
- **Lemosoda *40 centesimi*:** I agree to purchase ____ lemonsodas.
- **Chinotto *40 centesimi*:** I agree to purchase ____ Chinotto.
- **Crostini *35 centesimi*:** I agree to purchase ____ crostini.
- **Salad (lettuce, rucola, tomato, raw ham, salt and a teaspoon of oil served with bread) *1,50* €:** I agree to purchase ____ salads.
- **Fruit *40 centesimi*:** I agree to purchase ____ pieces of fruit.
- **Yoghurt low fat *60 centesimi*:** I agree to purchase ____ cups of yoghurt.
- **Crisps *35 centesimi*:** I agree to purchase ____ crisps.

- **Cucciolone *90 centesimi*:** I agree to purchase ____ cucciolone
- **Magnum Nuts *90 centesimi*:** I agree to purchase ____ Magnum Nuts
- **Cornetto *1* €:** I agree to purchase ____ Cornetto
- **Caramelle *60 centesimi*:** I agree to purchase ____ caramelle.

Non-food items: you will be charged by the minute; write the total number of minutes you want (or "0").

- **Textbook reading *1 centesimo per minute*:** I agree to purchase ____ minutes of textbook reading.
- **Research on the catalogue of the departmental library *1 centesimo per minute*:** I agree to purchase ____ minutes of research on the catalogue of the departmental library.
- **Research on the electronic catalogue of the university library *1 centesimo per minute*:** I agree to purchase ____ minutes of research on the electronic catalogue of the university library.
- **Weights and gym equipment for developing pectoral muscles *1 centesimo per minute*:** I agree to purchase ____ minutes of use of the weights and gym equipment.
- **Magazines (Panorama, l'Espresso, Economy) *1 centesimo per minute*:** I agree to purchase ____ minutes of magazine reading.
- **Video games *1 centesimo per minute*:** I agree to purchase ____ minutes of video game playing.
- **Music on the Virgin Radio, Italian Radio web sites, and CDs *1 centesimo per minute*:** I agree to purchase ____ minutes of web music.

From 12.45 to 17.30 you will be able to use the goods and activities that you agreed to purchase. At 17.30 you will be given a questionnaire. After completing the questionnaire, you will be paid 50 € minus what you spent on goods and activities.

II Instructions for the Experiment

This experiment lasts 5 h. During the first 15 min we will explain the instructions and show you the goods and activities that can be purchased. After this, you will have 4 h and 45 min to purchase goods and activities.

To purchase a good or an activity, simply tell the experimenter that you want something. During the experiment, participants can speak only with the experimenter and her assistants. Please do not talk to other participants.

The prices for goods and activities are listed below. If you order something (like a cup of espressino), it will be delivered to you as soon as possible.

We first list food items from the bar at the Faculty of Economics.

- **Espressino coffee *24 centesimi*.**
- **Espresso coffee *23 centesimi*.**
- **Chocolate Snack *50 centesimi*.**

- **Focaccia** ***90 centesimi***.
- **Sandwich** ***90 centesimi***.
- **Oransoda** ***40 centesimi***.
- **Lemosoda** ***40 centesimi***.
- **Chinotto** ***40 centesimi***.
- **Crostini** ***35 centesimi***.
- **Salad (lettuce, rucola, tomato, raw ham, salt and a teaspoon of oil served with bread)** ***1,50 €***.
- **Fruit** ***40 centesimi***.
- **Yoghurt low fat** ***60 centesimi***.
- **Crisps** ***35 centesimi***.
- **Cucciolone** ***90 centesimi***.
- **Magnum Nuts** ***90 centesimi***.
- **Cornetto** ***1 €***.
- **Caramelle** ***60 centesimi***.

Non-food items: you will be charged by the minute.

- **Textbook reading** ***1 centesimo per minute***.
- **Research on the catalogue of the departmental library** ***1 centesimo per minute***.
- **Research on the electronic catalogue of the university library** ***1 centesimo per minute***.
- **Weights and gym equipment for developing pectoral muscles** ***1 centesimo per minute***.
- **Magazines (Panorama, l'Espresso, Economy)** ***1 centesimo per minute***.
- **Video games on the RealArcade web site** ***1 centesimo per minute***.
- **Music on the Virgin Radio, Italian Radio web sites, and CDs** ***1 centesimo per minute***.

From 12.45 to 17.30 you will be able to use the goods and activities above. At 17.30 you will be given a questionnaire. After completing the questionnaire, you will be paid 50 € minus what you spent on goods and activities.

Debriefing Questionnaire

Please answer the following questions:
How did you hear about the experiment?

____

____

Are you a student at Bari? Yes ____ No ____. If you are not a student at Bari, what is your current employment status? __________.

If you are a student, what is your field of study: __________.

How many Euros do you spend per year on food? __________.

What is your age: __________.

Are you married? Yes ____ No ____.
What is your height: ____ meters.
Are you male or female: Male ____ Female ____.
What is your weight: ____ kilograms.
Were any parts of the experimental instructions confusing? Yes ____ No ____. If you answered yes, please describe which part was confusing:

____

____

What do you think this experiment was about? Please explain below:

____

____

Please rate the following goods and activities. Give each good or activity two ratings.

Rating 1: Immediate pleasure.

Does this good or activity give you immediate pleasure?

Is it fun, tasty, or pleasurable?

Circle one of three responses:

None immediate pleasure/Low immediate pleasure/High immediate pleasure

Rating 2: Delayed benefits.

Does this good or activity have benefits that last a few days?

Is it healthy or educational?

Circle one of three responses:

None delayed benefits/Low delayed benefits/High delayed benefits

	Rating 1: Immediate pleasure?	Rating 2: Long-term benefits?
Espressino coffee:	None/low/high	None/low/high
Espresso coffee:	None/low/high	None/low/high
Chocolate snack:	None/low/high	None/low/high
Focaccia:	None/low/high	None/low/high
Sandwich:	None/low/high	None/low/high
Oransoda:	None/low/high	None/low/high
Lemonsoda:	None/low/high	None/low/high
Chinotto:	None/low/high	None/low/high
Crostini:	None/low/high	None/low/high
Salad:	None/low/high	None/low/high
Fruit:	None/low/high	None/low/high
Yoghurt low fat:	None/low/high	None/low/high
Crisps:	None/low/high	None/low/high
Cucciolone:	None/low/high	None/low/high
Magnum nuts:	None/low/high	None/low/high
Cornetto:	None/low/high	None/low/high
Caramelle:	None/low/high	None/low/high
Textbook reading:	None/low/high	None/low/high
Library research:	None/low/high	None/low/high

Using gym equipment:	None/low/high	None/low/high
Panorama Magazine:	None/low/high	None/low/high
l'Espresso Magazine:	None/low/high	None/low/high
Economy Magazine:	None/low/high	None/low/high
Video games:	None/low/high	None/low/high
Music on the web:	None/low/high	None/low/high

Debriefing Statement

This was an experiment on consumer choice. We were interested in studying how people allocate their time among different activities.

Thank you very much for participating. Please feel free to contact me at e.carbone@dse.uniba.it at any time with questions, or if you would like to receive a copy of the research paper that results from this study. Thanks.

____

Enrica Carbone

P.S. Feel free to encourage your friends to participate in this study. However, we ask that you not tell them any details about the actual tasks involved. We want them to come without pre-formed expectations. Thanks.

Appendix 2

The Implications of Alternative Mappings from the Indices to the Weightings

Scenario 1: Consider the following mappings:

Temptation coefficients				Investment coefficients		
1/2	0	1	2	0	1	2
0	0.000	0.000	0.000	0.000	0.125	0.250
1	0.125	0.250	0.500	0.000	0.250	0.500
2	0.250	0.500	1.000	0.000	0.500	1.000

With these we get the following weighted expenditures:

	Temptation	Investment
Spot	58	42
Forward	59	41

Scenario 2: Consider the following mappings (note that expenditure is always allocated *x%* to temptation and *(100–x)%* to investment except when the good is rated zero on both indices):

Temptation coefficients				Investment coefficients		
1/2	0	1	2	0	1	2
0	0.000	0.000	0.000	0.000	1.000	1.000
1	1.000	0.500	0.333	0.000	0.500	0.667
2	1.000	0.667	0.500	0.000	0.333	0.500

With these we get the following weighted expenditures:

	Temptation	Investment
Spot	62	38
Forward	65	35

Acknowledgements The author thanks David Laibson for stimulation and discussions.

References

Carbone, E. (2006). Understanding intertemporal choices. *Applied Economics, 38*(8), 889–898

Harrison, G. W., Lau, M. I., & Melonie, W. B. (2002). Estimating individual discount rates in Denmark: A field experiment. *The American Economic Review, 92*(5), 1606–1617

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics, 112*(2), 443–477

Phelps, E. S., & Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *Review of Economic Studies, 35*(2), 165–180

Read, D., Loewenstein, G., & Kalyanaraman, S. (1999). Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic. *Journal of Behavioral Decision Making, 12*(4), 257–273

Read, D., & Van Leeuwen, B. (1998). Predicting hunger: The effects of appetite and delay on choice. *Organizational Behavior and Human Decision Processes, 76*(2), 189–205

Monty Hall's Three Doors for Dummies

A. Morone and A. Fiore(✉)

Keywords: Learning; Anomaly; Individual decision making; Experiment; Status quo bias

1 Introduction

Since its appearance in the literature, Monty Hall's three doors "anomaly" has attracted the attention of scientists from different fields: economists (Friedman, 1998; Nalebuff, 1987; Page, 1998; Palacios-Huerta, 2003; Slembeck & Tyran, 2004), statisticians (Morgan, Chaganty, Dahiya & Doviak, 1991; Puza, Pitt & O'Neill, 2005), psychologists (Granberg & Brown, 1995; Krauss & Wang, 2003) among others. The reason for such attention may be explained by the fact that this "anomaly" relies on a simple, even if counterintuitive, problem. There are three doors, and only behind one of them there is a big prize. At the beginning of the game, the player is asked to choose just one door. After that, an empty door among the not chosen doors is opened and the player is asked to make a new decision: either to stick with the first chosen door or to change and choose the remaining not-opened door. If people performed the Bayes' updating correctly, they should realise that switching is the best strategy because it doubles the percentages of winning (as we show in the next section). Nevertheless, the stylised fact from the American TV programme in which this game was firstly performed, and from the controlled experiments that replicated its basic structure (Friedman; Page; Palacios-Huerta; Slembeck & Tyran) is that only a low percentage of people choose to switch.

The aim of this paper is to contribute to understanding the possible reasons for which the people fail to adopt the best strategy, even in an environment in which

A. Fiore
Department of Economics, University of Bari. Via Camillo Rosalba 53, 70124 Bari, Italy
e-mail: afiore@dse.uniba.it

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.

© Springer-Verlag Berlin Heidelberg 2008

they perform the task repeatedly. Our experimental results drive us to conclude that the most common reason for this anomaly, that is, the misapplication of Bayes' rule, even if it contributes undoubtedly in reducing the anomaly, is not the only nor the most important reason for it.

The paper is structured as follows. In Sects. 2 and 3, we will introduce the problem and we will present the new experimental design, respectively. In Sect. 4 our results will be presented. In Sect. 5 some possible explanations will be proposed. Finally, in Sect. 6, we will draw some conclusions.

2 The Problem

Monty Hall's three doors is a particularly simple game. First, subjects are asked to choose one of three doors, that are equally likely to hide a big prize. Consequently, the first chosen door has a probability of one third, whereas the two left doors taken together have a probability of two thirds hiding the prize. Moreover, if we consider the remaining two doors, we know that with probability 1, one of them is surely empty. Then, when one of the two left doors is opened, knowing precisely which one is empty does not add any relevant information, and does not affect the probability that the first chosen door hides the prize or the probability that the prize is behind the not chosen pair. Nevertheless, it seems that the players are generally unable to recognize this. This is one of the most crucial points in the Monty Hall's problem, since it is directly related to the issue in decision-making concerning the manner in which people process new information and update beliefs. It is well-known that in the Monty Hall game the optimal strategy is to switch. This follows from a direct application of Bayes' rule: let us label the three doors A, B and C and assume a subject chooses door A. Additionally, Monty opens door B (that is an empty door, and the subject knows the door will be opened is empty). Now we can calculate the probability of winning by switching to C given that Monty opened B and the probability of winning by not switching to C given that Monty opened B:

$$\Pr(prize\,in\,C|Monty\,opened\,B) = \frac{\Pr(Monty\,opened\,B|prize\,in\,C)\Pr(prize\,in\,C)}{\Pr(Monty\,opened\,B)}$$
$$= \frac{(1)(1/3)}{(1)(1/3)+(1/2)(1/3)} = \frac{2}{3} \tag{1}$$
$$\Pr(prize\,in\,A|Monty\,opened\,B) = \frac{\Pr(Monty\,opened\,B|prize\,in\,A)\Pr(prize\,in\,A)}{\Pr(Monty\,opened\,B)}$$
$$= \frac{(1/2)(1/3)}{(1)(1/3)+(1/2)(1/3)} = \frac{1}{3} \tag{2}$$

A rational subject should be able to perform this calculation and therefore he/she should choose always to switch. Unfortunately, this is not the case, since many subjects decide to stay with their first choice.

Several possible reasons have been suggested to explain this "anomaly", for example: (1) subjects could make mistakes in updating the relevant probabilities, i.e., they do not perform the Bayes' rule correctly; (2) they could suffer of the *endowment effect* (Kahneman, Knetsch & Thaler, 1991; Knetsch, 1989) which captures the overvaluation of the winning probability of the owned door and/or the *status quo bias* (Samuelson & Zeckhauser, 1988) which is the preference to remain at the current door; (3) subjects could believe erroneously that the task, in addition to chance, entails some kinds of skill (in the psychological literature, this belief is called *illusion of control*) and, therefore, they could try to "guess" somehow the winning door using some skill as insight, and seeing one of the two remaining doors empty is nothing but a reinforcement of their prior belief; (4) subjects could act following as a strategy the *probability matching behaviour*, i.e., they could decide to choose not the optimal strategy (always switching), but they could choose each strategy according to their relative likelihood of success. In our case, this means that they should choose the two strategies, switching and not switching, in proportion of one third and two thirds, respectively.

Our experimental design enables us to discriminate better among these different explanations. In this perspective, our approach is different, inasmuch we developed a more radical "debiasing test" (Conlisk, 1996) compared to all previous experimental attempts, whose focus has been almost exclusively on some particular aspects that could be able to mitigate the "anomaly" and help people to behave rationally. These treatments were designed to endorse learning and to test other institutions recognized sensitive to anomalous choice behaviour.

3 The New Experimental Design

When we first approached this problem, we were puzzled, and we were moved by a simple idea: "if something is true there should be an easy way to explain it." Obviously, if it is true that "switching the door" is better than "keeping the door", 2/3 chance of winning is better than 1/3, then there should be an easy, understandable, and convincing way to demonstrate it. In order to determine this easy way, we modified the *Monty Hall's three doors* into the *Monty Hall's three doors for dummies*. As in the basic structure, there are three doors, and only behind one of them there is a big prize. At the start, as usual, subjects are asked to choose one of the three doors, but then they should be asked whether they would like to change the door they firstly chose with both the other two doors. In this way, they should readily realize that we were trading off 2/3 change of winning for 1/3, and they should take advantage of such opportunity promptly. As a consequence, the "anomaly" should disappear.

Consequently, we decided to run an experiment composed of two treatments: CONTROL, and FOR DUMMIES. In the CONTROL treatment, we substantially replicated Friedman' first treatment. Indeed, in this treatment participant was firstly asked to choose a door among three. Then, an empty unchosen door was opened.

Finally, subject was asked to stick with to the first chosen door or to switch to the remaining unchosen door. Conversely, in the FOR DUMMIES treatment, subjects chose a door, and then they were asked to change the chosen door with both the remaining two doors. In this case, no empty door was revealed. In order to make our results robust for monotonicity, we also ran a third treatment, the INTERMEDIATE treatment. In this case, the only difference was that, after an empty door one was opened, they were asked whether they wanted to keep the chosen door or whether they wanted both the other two doors (one closed and one open).

At this point, it should be worth summarizing the main features and the options available to subjects in each single treatment:

- CONTROL: once an empty door is opened, new information enters in the game and some probabilities have to be updated. All subjects should update their belief (i.e. the probability that the firstly chosen door hides the prize was 1/3 and it remains 1/3; the probability that the open door hides the prize was 1/3 and it fell to 0; the probability that the last door hides the prize was 1/3 and it rises to 2/3). In this treatment, it is easy to fail processing correctly the new information, subjects wrongly attached to the two still closed doors a 50–50 chance to hide the prize. In this sense, they fail the Bayesian updating
- INTERMEDIATE: once an empty door is opened, new information enters in the game once again and some probabilities have to be updated as well. In this treatment, since the two left doors are offered coupled, performing the correct Bayesian updating is easier, but it is still possible to fail it and attaching to the two closed doors a 50–50 chance to hide the prize
- FOR DUMMIES: after the first choice, no door is opened, and so there is no new information entering the game. The Bayesian updating is not needed and so it should be impossible to fail it

Otherwise, the three games were identical: all the relevant probabilities remained unchanged, but if in the CONTROL treatment subjects may fail to consider the opened and the left door as a pair, in the remaining treatments, they cannot fail to consider them a pair, because we actually offered them in pairs.

In this way, this simple design should allow us to establish whether subjects fail the Bayesian updating, or if they may present some psychological underpinnings that drive them in keeping the first choice, seeming to be irrational.

4 The Experiment

This paragraph is structured as follows. Section 4.1 illustrates design and procedures, while in Sect. 4.2 we present the main results. Finally, in Sect. 4.3 is provided an econometric analysis of determinants of learning.

4.1 Design and Procedures

The experiment was programmed using the Z-tree software (Fischbacher, 2007) and was run at the laboratory of ESSE (Economia Sperimentale al Sud d'Europa) of the University of Bari on January 2005.

Each treatment, lasting for about 45 min, was made up of 12 periods, 2 of which were trial periods.[1] The trial periods were necessary in order to acquaint subjects with both the task and the computerized interface. Before the experiment was started they got the chance to ask questions about the experiment's instructions. We paid particular attention in writing the instructions and in avoiding any possible misunderstanding and/or deception (instructions are available on request). For example, in explaining the structure of the game, we did not refer to any among the three cards as the opened one. This aspect is on trial in this statement: "Although, semantically, Door 3 [. . .] is named merely as an example (Monty Hall opens another door, *say*, number 3), most participants take the opening of Door 3 for granted and base their reasoning on this fact" (Krauss & Wang, 2003).

In each treatment, we had $N = 20$ subjects (*between-subjects* design), randomly assigned to the three treatments, all of them sat next to a PC terminal. The subjects could not see each other or communicate with the other subjects. Almost all of them were undergraduate students in Economics not familiar with previous similar experiments.

In the experiment the task was very simple: to pick on the screen a door among three, simply pressing a button. For each period the programme established which door hid the prize, and the subjects knew this.

This constituted the first stage of the game, the same in all the treatments. Then a new stage began. In all the treatments, information about subject's previous choice was displayed. Then, the three treatments differed. In the CONTROL treatment, the programme chose and showed an empty door to the subject and then asked him/her whether he/she wanted to keep his/her first choice or if he/she preferred to go for the remaining door. In the INTERMEDIATE treatment, the programme chose and showed an empty door to the subject and then asked whether he/she wanted to keep his/her first choice or if he/she preferred the un-chosen doors (i.e. one door is opened and visibly empty, the other one is still closed). Finally, in the FOR DUMMIES treatment, subjects were given the opportunity to change the chosen door with the other two doors. In the final stage, subjects were informed about their chosen door(s), the right option and about their pay-off. They gained 0.5 € whenever they chose the lucky door, and zero otherwise The average payoff, earned in half an hour, was 2.6 €.

4.2 The Experimental Results

Figure 1 summarizes the experimental results. It reports the switch rates in the three treatments over the ten periods. Comparing our CONTROL treatment with the experiments in the previous literature, we report an overall switch rate of 41.5%,

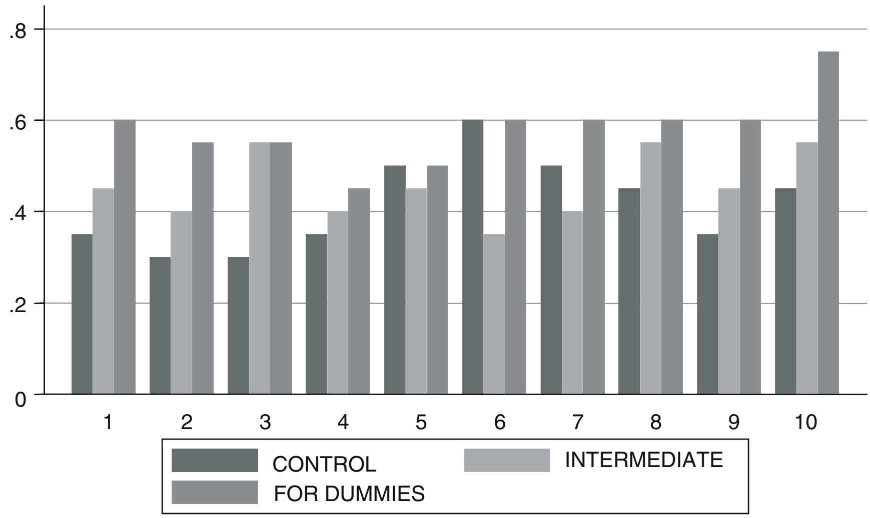


Fig. 1 Switch Rates by treatment and period

higher than in Friedman (28.7%), even if in line with results in other previous experimental studies (Page, 1998; Palacios-Huerta, 2003; Slembeck & Tyran, 2004). The overall switch rate in INTERMEDIATE and FOR DUMMIES is 45.5 and 58%, respectively. Whereas there is no statistical difference between the CONTROL and the INTERMEDIATE treatment (Wilcoxon rank-sum test, p-value = 0.4203), the FOR DUMMIES treatment is statistically different from both of them (Wilcoxon rank-sum test CONTROL/FOR DUMMIES, p-value = 0.0010; INTERMEDIATE/FOR DUMMIES, p-value = 0.0125). Therefore, we can conjecture that the FOR DUMMIES treatment could have turned out to be effective in shaping a different kind of behaviour, whereas the other two treatments are statistically indistinct.

Even though we observe a higher (or, at least, not smaller) switch rate in any single period under the FOR DUMMIES treatment, and a monotonic increasing pattern across the three treatments as expected, nevertheless the switch rate in the new framework is still too low, even in the last period (when it reaches its maximum at 75%). Indeed, if the real reason for this problem were the misapplication of the Bayes' law, since the new framework did not require subjects to make any probability updating, we should have observed a switch rate not statistically different from 100%. At this point, we could not consider the Bayesian updating failure as the leading explanation. We should look for other plausible explanations.

Before going in to details, we show in Fig. 2 the categorization of our experimental subjects according to their different switch rate during the experiment.

As can be seen, even though the percentage of completely rational subjects is undoubtedly higher under the for dummies treatment, i.e., 35% of subjects always switched, compared to the 5% in both other two treatments, nevertheless even in this case a not negligible percentage of subjects never switched (3 out of 20). We now try to list some of the possible explanations of this behavioural pattern.

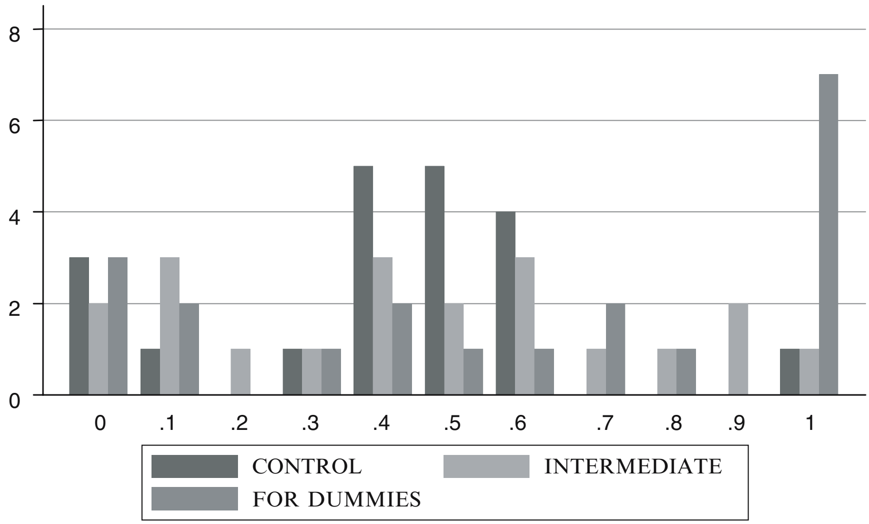


Fig. 2 Categorization of subjects according to their switch rates $N = 20$ in each treatment

As a first step, we test if the subjects behaved randomly. In order to test this hypothesis, we run a one-sided binomial test under the null hypothesis whereby switch rate is equal to 50% against the alternative hypothesis whereby the switch rate is greater than 50%. Very interestingly, we can reject the null hypothesis only for the FOR DUMMIES treatment (*p*-value = 0.0141), whereas we cannot reject the hypothesis of completely random behaviour in the CONTROL and INTERMEDIATE treatments (*p*-value = 0.9934 and *p*-value = 0.9105, respectively).

Given that probability matching behaviour has been invoked as a possible explanation for this "anomaly", as a second step, we investigate if the switch rate is significantly different from 2/3. We have already underlined that, according to this hypothesis, subjects would play strategies in relation to their likelihood of success (and switching has 2/3 chance of winning) and not the optimal strategy (in this case, they should always decide to switch), as the axiom of rationality would require. We can reject the hypothesis that subjects behaved according to the probability matching behaviour in all three treatments (CONTROL treatment, *p*-value = 0.0000; INTERMEDIATE treatment, *p*-value = 0.0000; FOR DUMMIES treatment, *p*-value = 0.0083).

4.3 Econometric Analysis

In this paragraph, we will follow Friedman (1998) and Slembeck & Tyran (2004) to estimate a simple learning model.

Following our predecessor, we have constructed a model that links the decision of switching or not with a set of determinants, as follows. The presence of learning is investigated by the use of the variable *Time* (the period number). In order to study reinforcement learning (or fictitious play; Erev & Roth, 1998), we use the *Switchbonus* variable (the cumulated earnings from always switching minus the earnings from always not switching); directional learning (or Cournot behaviour; Selten & Buchta, 1998) is studied by using the *Switchwon* variable (a dummy variable equal to 1 if in the most recent period the subject switched and won). We run a probit estimation introducing two more variables: $Time^2$, to test for concavity of learning, and *Switchlost*, another way to test directional learning (a dummy variable equal to 1 if in the most recent period the subject switched and lost). Results are reported in Table 1.[2]

Table 1 Maximum likelihood probit estimation

Dep. variable: *switch*	Friedman: first treatment	Friedman: second treatment	Slembeck–Tyran	Morone–Fiore
Constant	−1.090 (0.000)	-0.814 (0.000)	–	–
Time	0.055 (0.000)	0.032 (0.000)	0.0135 (0.013)	−0.0539 (0.109)
$Time^2$	–	–	−0.003 (0.008)	0.0044 (0.141)
Switchbonus	0.325 0.000	0.082 (0.000)	0.0017 (0.003)	0.0632 (0.004)
Switchwon	0.106 (0.344)	0.293 (0.000)	0.3323 (0.000)	0.3467 (0.000)
Switchlost	–	–	−0.2149 (0.000)	0.2735 (0.000)
Fiedman's treatments				
Intense	–	-0.243 (0.002)	–	–
Track	–	0.276 (0.009)	–	–
Advice	–	0.337 (0.012)	–	–
Compare	–	0.208 (0.069)	–	–
Slembeck–Tyran's treatments				
Competition	–	–	0.1435 (0.035)	–
Communication	–	–	0.1549 (0.022)	–
*Comp*Comm*	–	–	0.0468 (0.375)	–
Morone–Fiore's treatments				
Intermediate	–	–	–	0.0229 (0.661)
ForDummies	–	–	–	0.1211 (0.022)
Log likelihood	−589.9	−927.8	−956.88	−370.95
Pseudo R^2	–	–	0.2571	0.1073
NOBs	1.040	1.407	1.880	600

In Table 1, we report marginal effects, since coefficients derived from models of this sort have not the usual meaning and interpretation as in linear models, whereas marginal effects do have. In particular, the values of the computed marginal effects vary with the different values of the regressors. As regards the package we used in analysing our data (STATA), it reports the marginal effects at the sample means of the data for continuous variables, and for a discrete change from 0 to 1 for dummy variables.

Given these differences in interpreting coefficient estimates and marginal effects, our results are more easily comparable with those of Slembeck and Tyran (see footnote at Table 1). We can observe that among significant variables, *Switchwon* behaves quite exactly as in the previous analysis, whereas *Switchlost* has the same magnitude, but it goes in the opposite direction. That means that we cannot confirm precisely directional learning theory, because our results are not unambiguous: the variable *Switchwon* indicates that the probability of switching is 34.67% higher if the subject chose to switch and won in the most recent period, as we expected, instead *Switchlost* suggests that even if the subject chose to switch and lost in the preceding period, this fact increases the probability of switching by 27.35%. As regards the test for the other learning theory, reinforcement learning, the variable *Switchbonus* shows the expected direction and is more effective than in the previous analysis.

The negative effect for *Time* would suggest a downward trend to switch, on the contrary, the positive effect for $Time^2$ would show that we have a non linear and convex trend over time, but both of them are not significant.

Finally, as can be observed from the table, our last treatment, *ForDummies*, is significantly effective in increasing the probability of switching (by 12.11%).

In conclusion, our results are quite in line with previous ones, except for the fact that learning is not explained by the variable *Time* nor by directional learning (however, also in Friedman *Switchwon* is never significant) and that other determinants could contribute to explain the model. Rather, the optimal strategy is chosen only when sufficient favourable evidence has been accumulated.

5 Discussion

In this section, we summarize our main results and present some possible explanations for the behaviour observed in our experiment.

Considering the common explanations usually supported as likely candidates, already presented in Sect. 2, we can point out that:

(1) The *misapplication of Bayes' rule* is an important ingredient for this "anomaly", but we showed that it is not the driving explanation for this. Indeed, if it were the real motivation for this "anomaly", since in our FOR DUMMIES treatment no application of this rule was required, we should have observed no irrational behaviour under this treatment. Clearly, the sharp decline in the number of subjects that made an irrational choice across the three treatments may be attributed

to the fact that in the last experimental set up the game was simpler. Therefore, for a certain percentage of population, the problem lies in an incorrect probability updating. However, even in the last period, still 25% of people did not make the rational choice. This fact drives us to look for further explanations;

(2) Conversely, the *status quo bias* seems to have a not negligible role in understanding such a behaviour, at least for that fraction of people that never chose to switch. People seem to attach a higher value to their previous chosen door and, consequently, they seem to consider their already chosen door as their endowment. Indeed, very interestingly, even in the FOR DUMMIES treatment, at least 15% of subjects never decided to switch;

(3) The *illusion of control*, if on the one hand it could represent an explanation for the case in which subjects first face the game or when they undergo the game without repetition, it seems implausible in repeated experiments in which people have to opportunity to realize that actually the strategy of switching has a greater chance of winning with respect to the strategy of remaining (over all our treatments, only 38.71% of the choices of remaining were winners, whereas 65% of the choices of switching won);

(4) Finally, as regards the *probability matching behaviour*, we have already tested for this hypothesis, but our data reject it.

Additionally to the reason cited above, as a likely explanation, we can mention a kind of intertemporal inconsistency, i.e., subjects' decisions at the different decision nodes as if related to different "selves". Anyway, it is quite unlikely that people do the same "mistake" repeatedly over the ten periods: at the end, in our new treatment, they should realise that they are given the double chance of winning after the first stage.

6 Conclusion and a Final Remark

The main goal of our experiment has been to create a simple experimental set up that preserved all the basic features of the so called Monty Hall's problem, but such that subjects were not required to apply the Bayes' rule. This experimental design enabled us to discriminate among the most cited explanations for this kind of this problem.

Our main results have been clear: the misapplication of Bayesian updating is important in reducing the "anomaly", as we can derive from the monotonic increase of switch rate across the three treatments (the control that replicated exactly the structure of the game as in the TV programme and in the previous experiments, the second designed to test for monotonicity, but that is resulted not statistically distinct from the control, the new treatment in which subjects were simply required to choose between one door or two doors), but still, it does not appear as the leading explanation. In this sense, Monty Hall's three doors problem has proved to be stronger than commonly thought.

Having discarded some other common suggested explanations (the illusion of control, the probability matching behaviour), now we can affirm that this "anomaly", even if attenuated by design conditions, it is not a weak effect, but rather a systematic behavioural regularity. It could rely on some psychological underpinnings, such as the *status quo bias*. In this sense, a future line of research could be test for this effect. For example, a 'super for dummies' treatment could be implemented, in which subjects play only one stage in which they are required to choose among the possibility of one door or the possibility of two doors. In this case, the endowment effect would play no role. It is important to note that loss aversion, along with status quo bias, has proved to explain some other important phenomenon, such as the equity premium puzzle (for a review, see Camerer & Loewenstein, 2003). Moreover, our data suggest that also some learning models could be helpful in understanding such phenomena (indeed, our data support the reinforcement learning, so that switching would be chosen only when sufficient favourable evidence has been accumulated).

Finally, these results could contribute to dismiss the idea that people actually use probabilities at all in making some kinds of decisions. In this perspective, even the term "anomaly" may be not correct referred to the Monty Hall's problem and to other similar ones. Indeed, as we know, the neoclassical approach to economic science has on its own basis some rationality assumptions, and in particular, for handling probabilities, rational agents are assumed to use Bayes' rule, among other things. Consequently, whenever a behaviour is empirically found to depart from these rationality assumptions, the behaviour itself is labelled as "anomalous" (literally: "different, irregular", from ancient Greek: a- = no, homalòs = equal, regular) and on these anomalies a wide literature has flourished (for a survey, Thaler, 1991 and Camerer, 1995). However, if we find sufficient evidence that seems to challenge the rationality assumptions themselves, we should also revise our idea of what may be classified as an anomalous behaviour.

Notes

[1]We also run a regression to test whether trial periods have any significant effect in shaping behaviour in the following periods. We can provide details on request.

[2]The dependent variable is 1 in periods in which the subjects chose to switch and 0 otherwise. Friedman reports the coefficient estimates, considering a panel structure, whereas Slembeck and Tyran, and Morone and Fiore report the marginal effects after the probit estimation and they do not consider a panel structure.

References

Camerer, C. F. (1995). Individual decision making. In A. Roth & J. Kagel (Eds.), *Handbook of experimental economics* (pp. 587–703). Princeton: Princeton University Press

Camerer, C. F., & Loewenstein, G. (2003). Behavioral economics: Past, present, future. In C. F. Camerer, G. Loewenstein, & M. Rabin (Eds.), *Advances in behavioral economics* (pp. 3–51). Princeton: Princeton University Press

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature,* XXXIV, 669–700

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review, 88*(4), 848–881

Fischbacher, U. (2007) *z*-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Friedman, D. (1998). Monty hall's three doors: Construction and deconstruction of a choice anomaly. *American Economic Review, 88*(4), 933–946

Granberg, D., & Brown, T. A. (1995). The monty hall dilemma. *Personality and Social Psychology Bulletin, 21*(7), 711–723

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspective, 5*(1), 193–206

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifferene curves. *American Economic Review, 79*(5), 507–521

Krauss, S., & Wang, X. T. (2003). The psychology of the monty hall problem: Discovering psychological mechanisms for solving a tenacious brain tease. *Journal of Experimental Psychology, 132*(1), 3–22

Morgan, J. P., Chaganty, N. R., Dahiya, R. C., & Doviak, M. J. (1991). Let's make a deal: The player's dilemma. *American Statistician, 45*(4), 284–287

Nalebuff, B. (1987). Puzzles. *Journal of Economic Perspectives, 1*(1), 157–163

Page, S. E. (1998). Let's make a deal. *Economics Letters, 61*, 175–180

Palacios-Huerta, I. (2003). Learning to open monty hall's doors. *Experimental Economics, 6*, 235–251

Puza, B. D., Pitt, D. G. W., & O'Neill, T. J. (2005). The monty hall three doors problem. *Teaching Statistics*, *27*(1), 11

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1*(1), 7–59

Selten, R., & Buchta, J. (1998). Experimental sealed bid first price auctions with directly observed bid functions. In D. Budescu, I. Erev, & R. Zwick (Eds.), *Games and human behavior. Esssays in honor of amnon rapoport* (pp. 79–102). Mahwah, NJ: Erlbaum

Slembeck, T., & Tyran, J. R. (2004). Do institutions promote rationality? An experimental study of the three doors anomaly. *Journal of Economic Behavior & Organization, 54*, 337–350

Thaler, R. H. (1991). The winner's curse: Paradoxes and anomalies of economic life. New York: Free Press

Overconfidence in Predictions as an Effect of Desirability Bias

F. Giardini(✉), G. Coricelli, M. Joffily, and A. Sirigu

Keywords: Decision making · Overconfidence · Reward · Desirability · Accuracy

1 Introduction

Most people hold unrealistic positive beliefs about their personal skills, their knowledge (Fischoff, Slovic, & Lichtenstein, 1977), and their possibilities to overcome the performance of other individuals (Weinstein, 1980). This general tendency, called *overconfidence*, is a stable and pervasive finding both in many real-life domains and in several experimental settings. People are overconfident about their driving skills (Svenson, 1981), about their ability as basketball players (McGraw, Mellers, & Ritov, 2004), about their competence in financial and managerial problems (Camerer & Lovallo, 1999; Mahajan, 1992), and about their general knowledge (Juslin, 1994; Harvey, 1997). This systematic overestimation of one's own capabilities and probabilities of success can have important consequences, and sometimes results in suboptimal decisions.

While the existence of overconfidence is uncontroversial, its sources and determinants are still open to debate (Ayton & McClelland, 1997; Klayaman, Soll, Gonzalez-Vallejo, & Barlas, 1999).

In this study we contribute to this debate by demonstrating that overconfidence in predictions is related to the desirability of the predicted outcome. When people are required to forecast possible future events, they tend to be more confident in the occurrence of favourable events, with little or no regard for their objective likelihood.

F. Giardini
Institute of Cognitive Sciences and Technologies (ISTC), National Research Council, Via San Martino della Battaglia, 44, 00185, Rome, Italy
e-mail: francesca.giardini@istc.cnr.it

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*.
Theory and Decision Library C 42.

© Springer-Verlag Berlin Heidelberg 2008

We claim that forecasts are related to the desirability of the evaluated/predicted event, i.e. the more desirable the event, the stronger the belief that this event will happen. By manipulating the level of reward for the correct answer in a visual perceptual task we were able to highlight the presence and the principal characteristics of a *desirability bias*, which affects people's confidence. In our experiments we found a general increase in subjects' confidence levels under a reward versus a no-reward condition. Furthermore, the outcome desirability in terms of relative reward value biased subjects' confidence, leading them to believe that they were more accurate than they actually were.

In what follows we present two studies showing how a desirable result, i.e. a monetary reward, can bias people's confidence judgements in their perceptual accuracy, inducing them to be overconfident. Calibration studies have long investigated people's ability to match their judgements of the relative frequency of an event to the actual likelihood of that event. Perfect calibration occurs when average confidence is equal to the actual frequency of that event, and people are said to be "well calibrated". Unfortunately, this happens quite rarely. Several studies have shown that people are usually poorly calibrated, exhibiting either under- or overconfidence.

Overconfidence is the positive difference between mean reported confidence in the chosen answer and the percentage of correct answers (CONF – % correct answers >0). This phenomenon is preponderant in general knowledge or cognitive tasks (Brenner, Koheler, Liberman, & Tversky, 1996; Fischoff et al., 1977; Klayman et al., 1999; Koriat, Lichtenstein, & Fischhoff, 1980). The inverse phenomenon is underconfidence (CONF – % correct answers <0), which is more frequent in perceptual or in very easy cognitive tasks (Bjorkman, Juslin, & Winman, 1993; Juslin & Olsson, 1997; Keren, 1988).

Therefore, people usually overestimate their ability or knowledge in cognitive tasks, and underestimate the accuracy of their perceptions in sensory tasks.

1.1 Overview

This paper explores the effects of a desirable outcome on people's accuracy and confidence in a visual perceptual task. Effects of motivation on perception, judgement and decision making are well documented, but these effects usually refer to probability evaluation. For instance, the possibility of gaining money induces people to neglect or underestimate the base rate probabilities of events (Bar-Hillel, 1980; Kahneman & Tversky, 1996). Thus the perceived probability of a given event increases as a function of reward, even though the probabilities of success (base rates) are unchanged. A study by Ginossar & Trope (1987) suggested that goals may affect the use of base rate information, and there is some general evidence that motivation may affect the use of statistical heuristics. Generally speaking, the effects of goals and desires on reasoning, forecasting and memory are well documented (for a review, see Kunda, 1990), but less is known about how desirability affects people's confidence.

Here, we will investigate the effect of reward on people's confidence, that is, the degree of belief in a given hypothesis, judgement, or prediction. We claim that a desirable result makes people feel more confident in the possibility of getting it, compared with a neutral outcome, which is neither beneficial nor harmful to them. We call this phenomenon the *desirability bias* and we predict that it will induce individuals to be more confident when the possible reward is higher, all other things being equal. The desirability bias is a motivational effect working on the belief people hold about the likelihood of a certain outcome, and it should be independent from other effects, such as the difficulty of the task.

In Study 1 we tested the effect of three reward levels on confidence judgements in a perceptual task with fixed difficulty. In Study 2 we investigated how confidence judgements vary as a function of reward (low or high) for three levels of difficulty of the task. Manipulating the complexity of the task we induced three levels of accuracy: *Difficult* (Accuracy ≤ 0.5), *Intermediate* ($0.5 < \text{Accuracy} < 0.75$) and *Easy* (Accuracy > 0.75). Along with the reward groups we also tested one (Study 1) and three (Study 2) Control Groups, which performed the same task, with the same difficulty levels, but with no monetary incentives during the experiment. Control groups allow us to set base rate confidence and accuracy levels that are then compared with reward conditions.

We used a perceptual task in order to isolate the effect of motivation, and to exclude other possible explanations for overconfidence, such as failure to think of reasons why one might be wrong (Koriat et al., 1980) or individual's failure to assess the credibility or weight of the evidence (Griffin & Tversky, 1992). The task we implemented has three main characteristics. First, it is divided in two independent parts: the first part is constituted by a low-level perceptual task, requiring no reasoning and cognitive processing, while the second part requires an inferential process to evaluate the reward and to assess the confidence in the performance. Second, it makes it possible to directly correlate subjects' performance, i.e. accuracy of their responses, to reward and to variations in confidence, excluding any sort of other motivational effect, since the reward is displayed only when the perceptual part is over. Finally, the absence of feedback and the controlled number of trials allow us to rule out any kind of learning during the experiment.

2 Study 1

2.1 Participants

Twenty-seven undergraduate students (15 female and 12 male) were recruited to take part in a study at the Experimental Economics Laboratory (LabSi) of the University of Siena. All subjects were naïve with respect to the nature and aims of the experiment. Mean age of participants was 22 years (s.d. = 1.83).

2.2 Stimuli

We used a visual motion discrimination task typically used in neuro-physiological studies with monkeys (Britten, Shadlen, Newsome, & Movshon, 1992; Celebrini & Newsome, 1994). The stimulus display consisted of one white circle on a black background, containing 2000 black dots jiggling toward the right and the left side of the screen, with a fixed percentage of them coherently shifting toward either the left or the right (see Fig. 1).

The difficulty level was determined as the ratio between the velocity of the jiggling movement of the dots in the background and that of the linear movement toward one direction of the set that participants had to identify. We assessed this level in a previous pilot study, where we singled out a level of accuracy (i.e. percentage of correct answers) around 70%.

The background movement consisted of some dots jiggling in a random manner toward the right and some others toward the left, whereas the coherent set consisted of a fixed percentage of dots moving coherently toward only one direction.

Subjects had to identify the "coherent direction" of the dots, separating the coherently moving dots from the background movement. Each of the two directions was equally probable. The stimulus difficulty was set at the beginning of the experiment and then it was kept constant throughout the experimental session. Each stimulus was presented for 2 s.

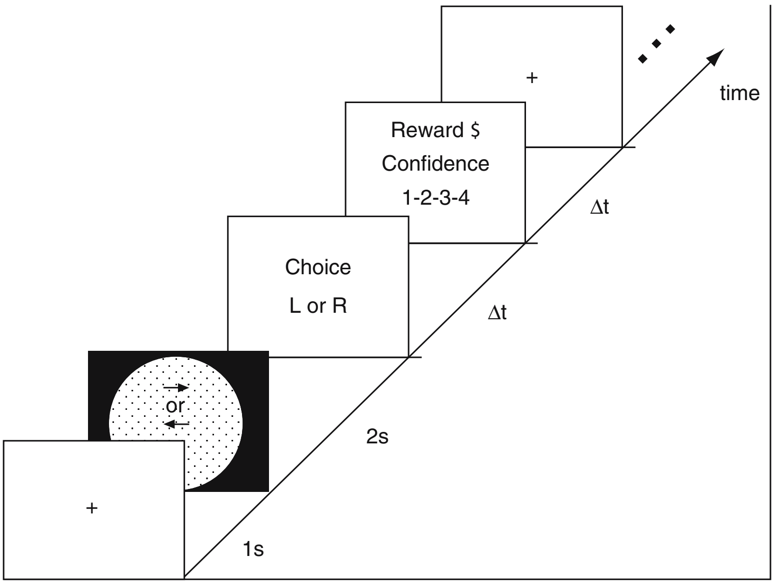


Fig. 1 Time course of the experiment in Study 1: fixation point (1 s), stimulus presentation (2 s), direction selection (left or right; self-paced), reward information (1, 5, or 15 €), and confidence scale (self-paced)

2.3 Procedure

Two groups of subjects were tested in two separate sessions. One group of eighteen subjects participated in a *reward condition* and the second group of nine subjects (control) took part in a *no-reward condition*. Participants had to discriminate the direction of moving dots showed on a computer screen placed in front of each of them.

Each trial began with a fixation point lasting 1 s. which directed the subject's attention toward the centre of the screen, where the stimulus was going to appear. The fixation point was followed by the stimulus presentation (2 s). At the end of the stimulus presentation, the participants saw on the screen the following question: "What was the dots' direction?", and below "Left or Right". They chose (self-paced choice) by pressing the corresponding arrow on the PC keyboard. Once one of the two arrows was pressed they could not modify their choice.

This was followed by a blank screen and then both the reward amount and a confidence scale appeared. All three amounts of reward (1, 5 and 15 €) were equally probable, and subjects were instructed that the computer program randomly paired rewards with stimuli. The uncertainty level for the stimulus recognition was always the same and no correlation existed among reward amounts and stimuli. This was explicitly stated in the Instructions and reminded to the subjects at the beginning of the experimental session.

The reward was showed in the upper part of the monitor (If you detected the correct direction you could be rewarded with ... Euro), while in the lower part appeared the question about the degree of confidence (How confident do you feel you detected the correct direction?). Confidence was measured on a 4-points confidence scale ranging from 1. "Not sure at all" to 4. "Really sure", with two intermediate values (2. "Not so sure" and 3. "Sure enough"). Subjects used left and right arrows on the keyboard to state their confidence and they could modify their choice until they pressed "Enter" to confirm it. There was no time limit for reporting confidence level.

Once they reported their confidence they pressed "Enter" to go to the next trial (Fig. 1). No feedback was provided to subjects, neither about the correct direction nor about their winnings. To summarize, the time course of the task was: [fixation point → stimulus → direction choice → possible reward → confidence judgement]→ [fixation point → (...)].

Subjects performed 57 trials, 9 of which were training trials aimed to get them familiarized with the task, while the remaining 48 were experimental trials. There was no time limit for completing the task. At the end of the session the subjects in the reward condition were paid accordingly to their performance in one trial randomly drawn by the computer out of the 48 trials (the 9 training trials were excluded). If their response in the drawn trial was correct they were paid accordingly to the reward shown during that trial, otherwise they only received the participation fee (3 €). These features of the experiment were properly explained in the Instructions subjects read before starting the session.

In a separate session, a Control Group, recruited following the usual procedure of the LabSi, was assigned (randomly) to the no-reward condition and performed an identical task (9 training trials and 48 experimental trials) except for the reward, which was neither mentioned nor displayed. Subjects in this condition were simply asked to individuate the direction and to assess their confidence on the 4-points scale, and received a participation fee of 3 €. Each experimental session lasted approximately 30 min.

2.4 Results

We found that confidence judgements of the correct answer vary with the amount of monetary reward (data from reward condition; repeated measures ANOVA, $F_{(2,17)} = 6.74$, $P = 0.0034$). Confidence increased as reward increased (as shown in Fig. 2), with significant differences between 1 and 15 € (Wilcoxon signed-rank test $z = -2.268$, $P = 0.0233$), and between 5 and 15 € (signed-rank test $z = -2.686$, $P = 0.0072$). No statistically significant difference was found between 1 and 5 € (signed-rank test $z = -0.982$, $P = 0.362$).

Regression analysis (Order Probit model, Table 1) can help us to understand where and how the reward enters the process.

We considered confidence (as reported in the confidence scale, i.e. takes values 1–4) as a function of reaction time (RT, equal to the response time of subject's choice of the direction of the moving dots), accuracy (A, equal to 1 if correct and 0 if incorrect), and reward level ($, only in reward condition, i.e. takes values of 1, 5 or 15). In the control group, which did not receive any reward, confidence level was a function of accuracy (A) and Reaction Time (RT). The time taken to respond as well as the accuracy of the responses determined subjects' confidence judgements when they did not have the possibility of getting any monetary reward

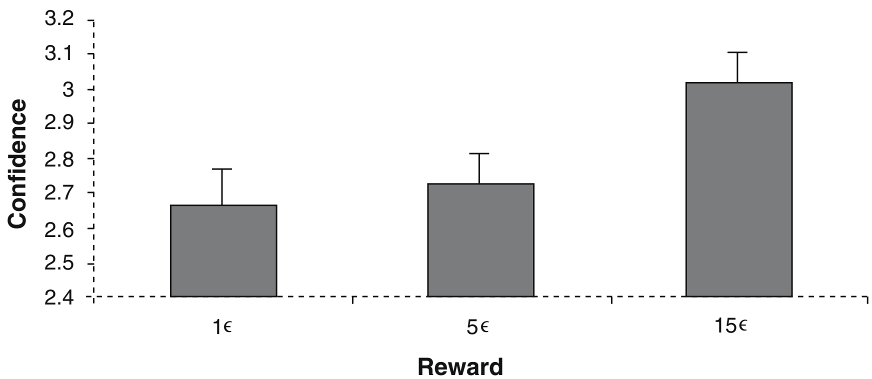


Fig. 2 Mean confidence (+standard errors) for the three different reward levels. Confidence was significantly higher for 15 compared with 1 and 5 €

Table 1 Study 1 regression analysis. Confidence levels as a function of Accuracy, reward level (only in reward condition) and reaction time. Regression analysis (Order Probit). The dependent variable Confidence takes values of 1–4; A = accuracy, equal to 1 if correct and 0 if incorrect; $ = reward level (i.e. 1, 5, or 15); RT = reaction time

Regression analysis order probit: dependent variable is "*CONFIDENCE*"

a. *Data from experimental sessions ($)*

Variable	Coeff.	Std. error	Z	P > |z|
A	0.133	0.082	1.64	0.102
$	0.035	0.006	5.51	0.000
RT	−0.0003	0.000	−7.67	0.000

Number of obs = 864
Log likelihood = -997.39701
Prob > Chi^2 = 0.0000

b. *Data from control sessions (no $)*

Variable	Coeff.	Std. error	Z	P > |z|
A	0.336	0.118	2.84	0.004
RT	−0.0002	0.000	−3.77	0.000

Number of obs = 432
Log likelihood = −500.13509 Prob > Chi^2 = 0.0000

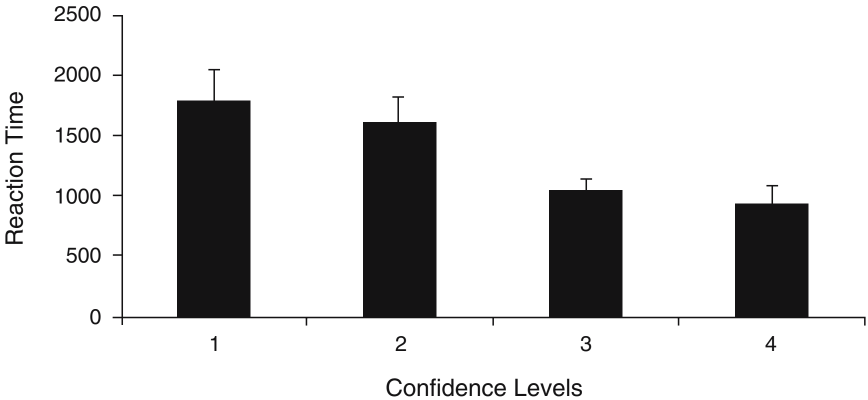


Fig. 3 Mean reaction time (RT; data collapsed over reward and no-reward conditions) for each level of confidence judgement of correct answer (with 1 – "Not sure at all" and 4 – "Really sure"). RT was inversely correlated with Confidence level

for a correct answer. By contrast, in the experimental condition, reward ($) and RT were significantly correlated with the confidence level, whereas accuracy was not. The inverse relationship between RT and confidence present in both conditions is shown in Fig. 3. The mean accuracy was not significantly different between the reward (Mean = 0.71) and the no-reward (Mean = 0.73) conditions (two-sample Wilcoxon rank-sum test $z = -0.258$, $P = 0.79$).

Thus, the presence of a monetary reward biases individuals' confidence, no matter how accurate they have been. That is, the possibility of receiving a large reward induced them to feel more confident.

The overall level of accuracy was constant in both conditions and during the whole experiment, thus this increase in confidence cannot be accounted for by a parallel increase in accuracy. Moreover, subjects were accurate in approximately 70% of the cases both in the experimental and in the control condition, with no appreciable changes in confidence for 1 and for 5 €, but with a significant increase in confidence for 15 € in reward trials.

In our analysis we considered actual accuracy as a proxy of the event probability of correctly performing the experimental task. Measuring over- and underconfidence we found a significant difference between results with the lower rewards and results obtained with the highest one (1 vs. 15 €, signed-rank test $z = -2.267, P = 0.0234$; 1 vs. 5 €, signed-rank test $z = -0.937, P = 0.35$; 5 vs. 15 €, signed-rank test $z = -2.68, P = 0.0073$).

Figure 4 shows over- and underconfidence for the three different reward levels. Overconfidence (CONF – % correct answers >0) appeared only with the highest reward (15 €), whereas underconfidence (CONF – % correct answers <0) was found for the two lowest rewards. Underconfidence was found also in the control group, in line with the results about confidence judgements in perceptual tasks reported in the literature (for a review, see Baranski & Petrusic, 1994).

Regarding the control group, we found an accuracy level (73% of correct responses) in line with the average difficulty of the task, and we also found underconfidence (CONF – % correct answers = −0.04), as predicted by theories of underconfidence in perceptual tasks.

These findings confirm our prediction that there exists a desirability bias, which overcomes accuracy and induces people to rely on a possible reward more than on actual accuracy.

Moreover, these results show the effect of relative reward on confidence judgements.

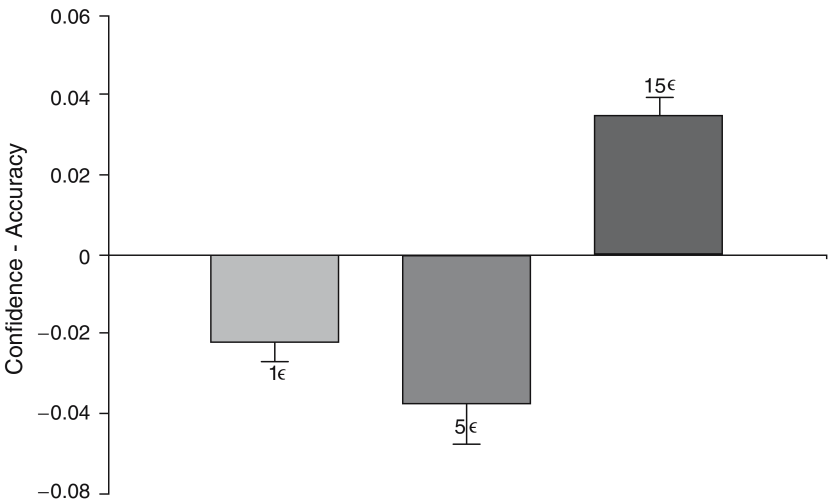


Fig. 4 Confidence (CONF/4) – Accuracy (%correct) for the three reward amounts. Results show underconfidence for 1 and 5 € and overconfidence for 15 €

3 Study

In this study, we investigated whether reward effect and desirability bias are present for other intervals of uncertainty. We tested subjects for three different difficulty levels (*Easy*, *Intermediate* and *Difficult*) and two rewards (2 and 10 €). We reduced the number of rewards, since in Study 1 we did not find any significant differences between the two lower rewards.

3.1 Participants

One hundred twenty-three undergraduate students (58 female and 65 male) from the University of Siena were recruited and randomly assigned to one of six groups. Three groups of subjects participated in reward conditions (25 for each condition, *Easy $*, *Intermediate $* and *Difficult $*) and the others three groups took part in no-reward conditions (16 control subjects for each condition, *Easy no$*, *Intermediate no$* and *Difficult no$*). All subjects were naïve with respect to the nature and aims of the experiment. Mean age of participants was 21.90 years (s.d. = 2.0096).

3.2 Stimuli

As in Study 1, we used a visual motion discrimination task (see Fig. 5). Three different difficulty levels were set in order to obtain three different levels of accuracy (in terms of percentage of correct responses). For subjects in the *Easy* condition we expected average accuracy to be higher than 0.75, for the *Intermediate* difficulty we expected results in the interval between 0.5 and 0.75, and for the *Difficult* condition we expected average accuracy to be lower than 0.50.

We assessed these conditions in a previous pilot study, where we singled out three levels of accuracy (i.e. percentage of correct answers) by manipulating the ratio between the velocity of the jiggling movement of the dots in the background, and that of the linear movement toward one out of four directions that participants had to identify (as it is in Study 1). The dots moved jiggling in a random manner toward one of four directions (right, left, up or down), whereas a fixed percentage of them moved coherently toward only one direction. We introduced two more directions (up and down) in order to increase the difficulty level.

Participants in the experiment had to identify the "coherent direction", individuating the coherently moving dots out of the background movement. The stimulus difficulty was set at the beginning of the experiment for each condition and then it was kept constant throughout the experiment. Each stimulus was presented for 2 s. The stimulus direction was randomized and controlled by the computer program, thus each of the four directions were equally probable and their single probability of occurrence was 25%.

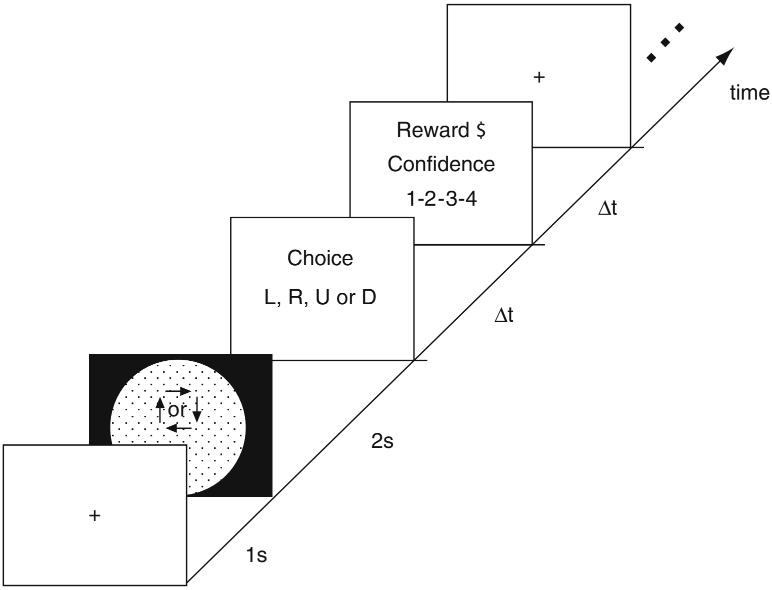


Fig. 5 Time course of the experiment in Study 2: fixation point (1 s), stimulus presentation (2 s), direction selection (up, down, left or right; self-paced), reward information (2 or 10 €), and confidence scale (self-paced)

3.3 Procedure

The sequence of events and the time course of the study was the same as in Study 1, thus [fixation point → stimulus → direction choice → possible reward → confidence judgement] → [fixation point → (...)].

The subjects were tested during six separate sessions and each and every subject participated in only one session.

Each trial began with a 1 s fixation point followed by the stimulus (2 s). After the stimulus presentation ended, participants saw on the screen the following question: "What was the direction of the dots? Left – Right – Up – Down". They responded by pressing the corresponding arrow on the PC keyboard (self-paced choice). Then, the screen was cleared and the reward and confidence scales appeared.

In the *reward conditions*, 2 and 10 € were equally probable, and subjects were instructed that the computer program randomly paired rewards with stimuli. The difficulty level for the stimulus recognition was always the same during the experiment and no correlation existed among reward amounts and stimuli. In order to avoid learning effects, no feedback was provided to participants. Subjects performed 72 trials (8 training and 64 experimental trials).

At the end of the session the subjects in the three *reward conditions* ($) completed a questionnaire and then they were paid accordingly to their performance in one trial randomly drawn by the computer out of the 64 trials (the eight training trials were excluded). If their response in the drawn trial was correct they were paid according to the reward showed during that trial, otherwise they received only the participation fee (3 €).

In three separate sessions, three control groups performed an identical task (8 training trials and 64 experimental trials) with the same three difficulty levels, except for the reward, which was neither mentioned nor displayed. These subjects were simply requested to individuate the direction and to assess their confidence on the 4-point scale (described above). At the end of the experiment they were asked to complete a questionnaire. Participants in the control groups received a show-up fee of 3 €. Each session lasted approximately 30 min.

3.4 Questionnaires

In this study we introduced a questionnaire to investigate the perceived difficulty and the determinants of subjects' confidence both in the reward and in the no-reward condition. In the former case the questionnaire was presented before subjects were informed about their winning, in order to avoid any kind of motivational or affective effects.

The questionnaire consisted of three questions regarding: difficulty (Question one: "According to you, the task was: (1) very easy; (2) fairly easy; (3) fairly difficult; (4) very difficult; (5) impossible"), accuracy (Question two: "According to you, what was the percentage of correct responses you gave?", subjects responded by circling the chosen percentage on a ten-point scale); and confidence (Question three: "According to you, which of these elements determined your confidence judgement?" – "(a) the perception of the stimulus; (b) the time required to make your choice; (c) the amount of the possible win; d. the perception of the stimulus and the amount of the possible win"). The questionnaire for the control groups was identical except for Question three, where any reference to reward was excluded (Question three: "According to you, which of these elements determined your confidence judgement?" – "(a) the perception of the stimulus; (b) the time required to make your choice").

The rationale for introducing questionnaires was the need to compare the 'trial by trial' evaluation (significantly and unequivocally affected by the displayed rewards), with the global estimate of difficulty, perceived accuracy and confidence. In other words, we were interested in assessing whether the participants, at least at the end of the task, were aware of the desirability bias. Moreover, questionnaires provided a subjective evaluation of the objective accuracy participants achieved.

3.5 Results

This study confirmed the results of Study 1, showing that the confidence level for a correct response varied with different reward levels (2 or 10 €) (data from reward condition; repeated measures ANOVA, $F(1, 74) = 50.15$, $P < 0.00001$).

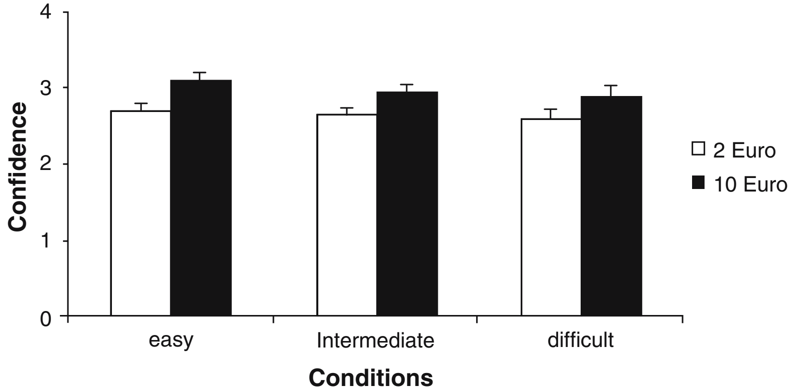


Fig. 6 Mean confidence level (+standard errors) as a function of the possible reward (2 or 10 €) for each condition (easy, intermediate, and difficult). Confidence increased as reward increased in each of the three conditions

Figure 6 shows how confidence was significantly higher for 10 € with respect to 2 €, in each condition (*Easy*, Wilcoxon signed-rank test z = 3.396, $P = 0.0007$; *Intermediate*, signed-rank test z = 4.315, $P < 0.0001$; and *Difficult*, signed-rank test z = 4.112, $P < 0.0001$).

The mean accuracy level was 0.84 (SD = 0.055), 0.61 (SD = 0.068) and 0.38 (SD = 0.073), for easy, intermediate and difficult condition, respectively. Thus, accuracy was significantly different for the three different difficulty levels (Kruskal–Wallis $Chi^2(2)$ = 63.563, $P = 0.0001$, data from reward condition; and Kruskal–Wallis $Chi^2(2)$ = 24.239, $P = 0.0001$, data from no-reward condition). However, there was no significant difference between accuracy in reward and control conditions for each level of difficulty of the task (two-sample Wilcoxon rank-sum test z = 0.866, P = 0.38, for easy; z = −0.414, P = 0.68, for intermediate; and z = 0.161, P = 0.87, for difficult). We found again a significant inverse correlation between RT and confidence level (Fig. 7 data from all conditions).

The Regression analyses using the data from the reward conditions (Order Probit model, Table 2) show that for the easy and difficult conditions the confidence was a function of the accuracy, reward level, and reaction time (inversely related). The results from the intermediate condition show that confidence judgements depended only on reward level and on reaction time (inversely related).

Thus in this condition we found, as in Study 1, that rewards by-passed the effect of the actual accuracy and biased subjects' confidence level. Results from the control conditions confirm that without the presence of rewards the determinants of confidence judgements are always accuracy and reaction time (inversely related). Furthermore, this study demonstrated that the *desirability bias* remains stable for different levels of difficulty of the task.

Figure 8 shows the pattern of over and under-confidence for different levels of difficulty of the task. In the Easy condition we found underconfidence (CONF – % correct answers <0) for both levels of rewards (2 and 10 €); in the difficult

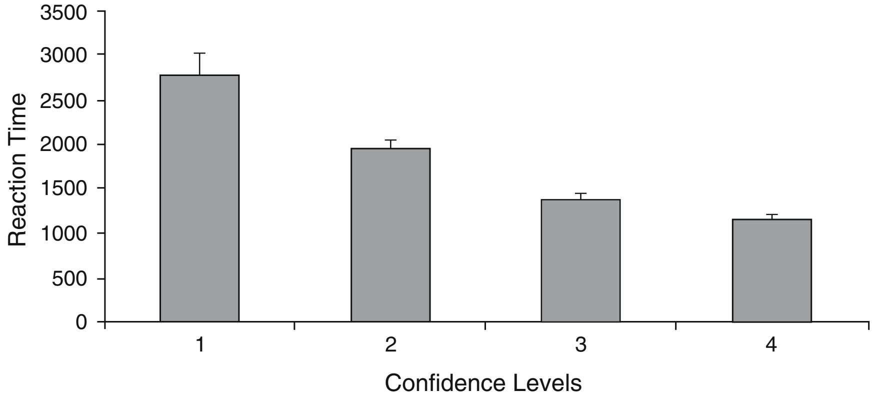


Fig. 7 Mean reaction time (RT; data collapsed over reward and no-reward conditions) for each level of confidence judgement of correct answer (with 1 – "Not sure at all" and 4 – "Really sure"). RT was inversely correlated with Confidence level, as in Study 1

Table 2 Study 2 regression analysis. Confidence levels as a function of Accuracy, reward level (only in reward condition) and reaction time. Regression analysis (Order Probit). The dependent variable Confidence takes values of 1–4. A = accuracy, equal to 1 if correct and 0 if incorrect; $ = reward level (2 or 10); *RT* reaction time

Regression analysis order probit: dependent variable is "*CONFIDENCE*"

a. *Data from experimental sessions with rewards for the three levels of difficulty (easy, intermediate, and difficult)*

	Easy				Intermediate				Difficult			
Variable	Coeff.	Std error	Z	P > |z|	Coeff.	Std error	Z	P > |z|	Coeff.	Std error	Z	P > |z|
A	0.2300	0.0646	3.56	0	0.0482	0.0501	0.96	0.336	0.1350	0.0552	2.44	0.015
$	0.0435	0.0069	6.3	0	0.0603	0.0063	9.62	0	0.0569	0.0068	8.33	0
RT	−0.0002	0.00002	−8.5	0	−0.0001	0.00001	−9.4	0	−0.0001	0.00002	−7.0	0

b. *Data from control sessions with no reward for the three levels of difficulty (easy, intermediate, and difficult)*

	Easy				Intermediate				Difficult			
Variable	Coeff.	Std error	Z	P > |z|	Coeff.	Std error	Z	P > |z|	Coeff.	Std error	Z	P > |z|
A	0.6687	0.0778	8.59	0	0.3388	0.0690	4.91	0	0.2749	0.0718	3.83	0
RT	−0.0004	0.00003	−10.7	0	−0.0001	0.00002	−4.9	0	−0.0002	0.00003	−8.7	0

condition the result was inverted, thus subjects were always overconfident (CONF – % correct answers >0); whereas in the intermediate condition we found overconfidence when the reward was 10 €, and approximately calibrated judgements for the cases in which the reward was 2 €. Note that the average accuracy in the intermediate condition was slightly lower (61%) compared to the average accuracy observed in Study 1 (71%). We observe (Fig. 9) underconfidence for the lowest reward (2 €) and overconfidence for the highest reward (10 €) for level of accuracy equal to 65%.

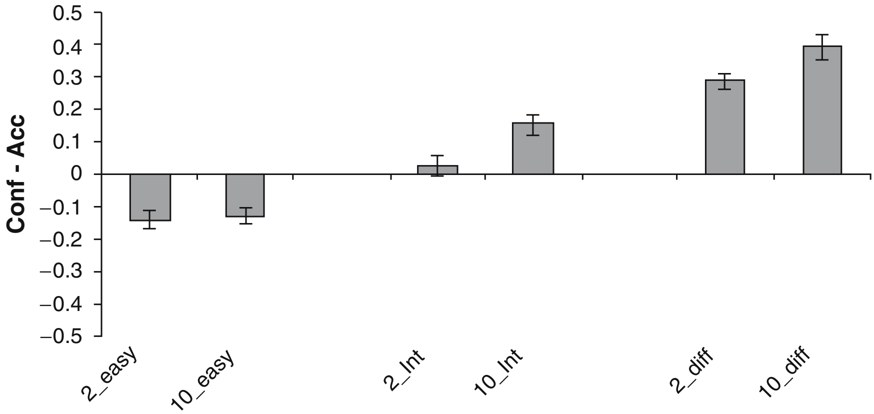


Fig. 8 Confidence (CONF/4) – Accuracy (%correct) for each difficulty level (easy, intermediate, and difficult) and for the two reward amounts (2 and 10 €)

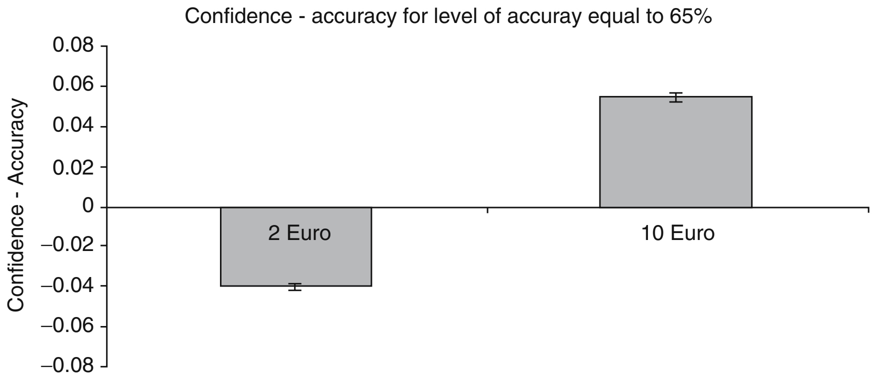


Fig. 9 Confidence (CONF/4) – Accuracy (%correct) for level of accuracy equal to 65%. As in study 1, we found overconfidence for the highest reward (10 €) and underconfidence for the lowest reward (2 €)

The analysis of the questionnaires allowed us to compare the trial-by-trial performances of subjects with their overall evaluation of their own choices. We were interested in checking whether subjects were aware or not of the role of reward and of the stimulus difficulty. On average, subjects considered the task quite difficult (mean difficulty evaluation = 3.43, SD = 0.94), with a significant difference in the relative frequencies of the responses (ranging from 1-very easy to 5-very difficult) for the three conditions (Question 1, $Chi^2(8) = 10.60$, $P = 0.22$). Figure 10 and b shows the discrepancy between Experimental Accuracy (the percentage of correct responses people gave) and the Reported Accuracy (the percentage of correct responses they thought they gave), in reward (Wilcoxon signed-rank test, $z = 7.526$, $P < 0.0001$) and no-reward conditions (Wilcoxon signed-rank test, $z = 5.56$, $P < 0.0001$) (Question 2). The difference between Experimental and

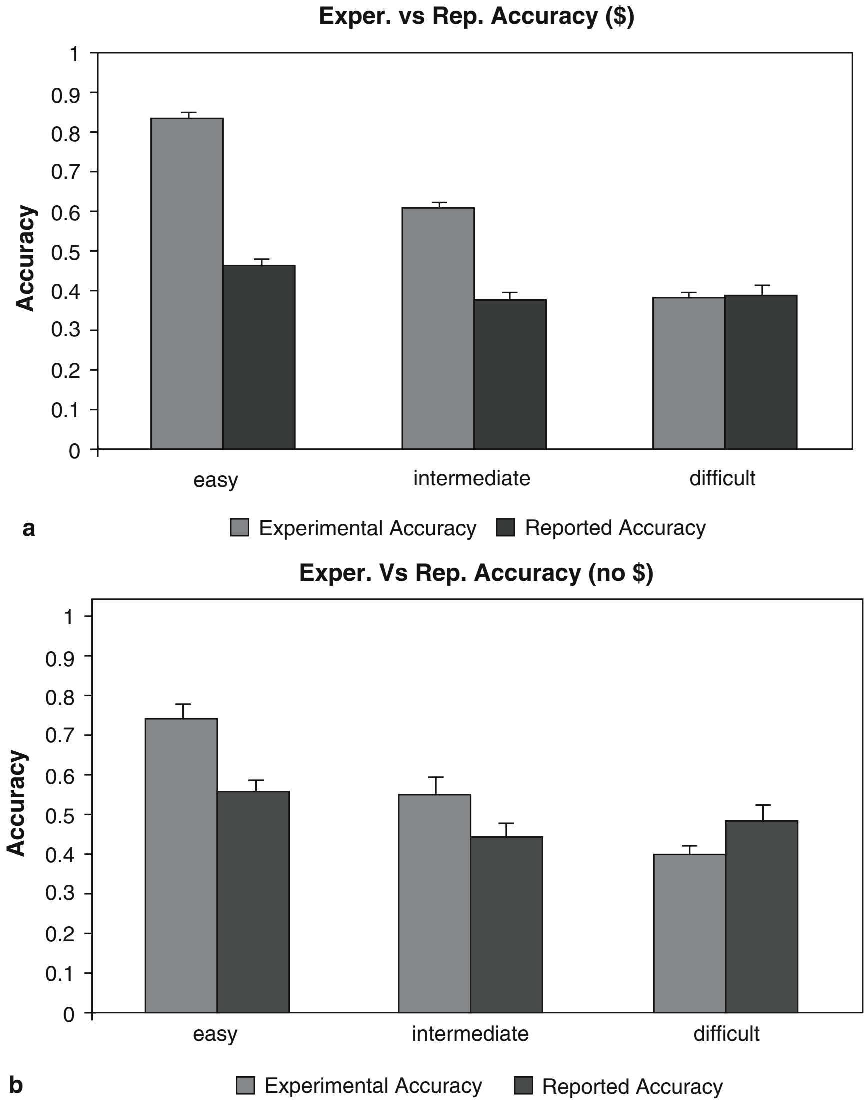


Fig. 10 **a** Comparison between Experimental Accuracy (percentage of correct responses during the task) and Reported Accuracy (estimated percentage of correct responses at the end of the task) shows that people underestimated their performances in all conditions. **b** Comparison between Experimental Accuracy (percentage of correct responses during the task) and Reported Accuracy (estimated percentage of correct responses at the end of the task) shows that people also underestimated their performances in the control condition, but this underestimation is absent in the difficult condition, where we instead found overestimation

Reported accuracy could be explained by the absence of immediate reward in the Reported Accuracy, so that people underestimated their performances, as usually happens in perceptual tasks.

The relative frequency of different reasons for confidence (Question 3: a, b, c, d) strictly depended on the difficulty level (difficult, intermediate, easy; $Chi^2(6) = 9.21$, $P = 0.16$). Almost all subjects (88%) in the Easy condition attributed their confidence to the stimulus perception, but this percentage decreases as uncertainty increased (66% intermediate, and 46% difficult). Moreover, around 30% of the subjects in the Difficult condition attributed their confidence to both the perception and the reward. This (ex-post) awareness did not prevent them from being biased by reward, as showed in Fig. 6. On the contrary, the control groups in all the three conditions attributed their confidence mainly to the stimulus perception (50% easy, 57% intermediate, 47% difficult).

4 Summary and Conclusions

People are often inaccurate in predicting their performances or their rates of success in many different domains, and many different explanations have been put forward. We suggest a general mechanism, which could work in a wide variety of domains and situations. Our findings indicate that people become relatively more confident about the occurrence of events associated with high rewards, compared with neutral events. These findings are in line with the theory of anticipatory representations by Miceli & Castelfranchi (2002), who proposed a theoretical account of expectations as a class of goal-driven anticipations.

We assume that the desirability of an outcome directly affects confidence in the occurrence of that outcome, inducing people to be more confident in it, when compared with a neutral or negative result. This assumption has been experimentally tested, and the results confirmed our hypothesis. Although the reward was merely possible, participants showed significant increases in their average confidence when a higher reward was presented. The correlation between reward and confidence was not linked to any appreciable change in accuracy, so we can reasonably conclude that the only factor modifying individuals' confidence in their choices was the reward. This means that people were not more accurate or faster in responding to the stimulus, they were just more confident in their performance when the possible reward was higher, compared to trials where the reward was lower.

Other studies (Bar-Hillel & Budescu, 1995; Irwin, 1953) tried to demonstrate the effect of a rewarding outcome on confidence levels, but motivation was not isolated from other variables, such as accuracy, so that they failed to detect any relationship between confidence and a desirable result.

By the contrary, our findings support the general hypothesis that the presence and the amount of a desirable outcome can affect people's confidence in their predictions. The pattern of confidence changes becomes especially striking when it turns into overconfidence for the highest reward in Study 1, and in the intermediate condition of Study 2 (when accuracy equal 65%). We assume that when the actual probability of the event to be predicted is extremely low or extremely high, the motivational aspects are less important in determining people's confidence judgements. Instead, when uncertainty is higher than chance but lower than certainty, judgements are desirability driven. This may happen because in this range participants were more sensitive to external (such as reward) or internal motivational cues that might drive their judgements. Considering the results of the questionnaire at the end of Study 2, we suggest that this phenomenon works at an unconscious level. Indeed, subjects indicated the perception of the stimulus as the main determinant of their confidence judgements, whereas they did not recognize the actual effect of reward.

The *desirability bias* affects people's confidence, inducing them to be more confident in the occurrence of a positive outcome, compared with a neutral one. Similar results have been reported in the psychological literature regarding "positive illusions" (Taylor & Brown, 1988), i.e. unrealistic positive beliefs about the self and one's own possibility of success and well-being. These illusions seem to be quite pervasive in human life. However, their causes are not entirely clear and the main question is whether they exert a positive or a negative influence on people's choices, behaviours and lives.

We predict that the desirability bias is a general phenomenon which could play a role in explaining optimistic overconfidence in predictions. People overestimate their possibility of achieving positive results because the "desirability bias" affects their confidence, causing them to believe that the desired result is more easily achievable. In other words, people do not simply expect events, but they actively desire positive outcomes, thus feeling more confident in the possibility of achieving the desired result.

This can be true also when the reward is not materially but psychologically relevant, such as self-esteem, social approval, and even cognitive dissonance reduction or avoidance (Blanton, Pelham, DeHart, & Carvallo, 2001; Festinger, 1957). This result is not trivial, and it could help preventing lots of mistakes due to overwhelming confidence in one's own capabilities and possibilities of success.

Acknowledgements We thank Cristiano Castelfranchi, Maria Miceli and Karen Reilly for helpful comments and suggestions. This work was funded by the PAR project 2003 "Anticipatory minds" – University of Siena (to FG); Action Concertée Incitative, Systemes Complexes from CNRS (to GC and AS); and Human Frontier Science Program (RGP 56/2005; to GC and AS). We are extremely grateful to Francesco Lomagistro for his assistance in recruiting subjects and running the experiments.

References

Ayton, P., & McClelland, A. G. R. (1997). How real is overconfidence? *Journal of Behavioral Decision Making, 10*, 279–285

Baranski, J. V., & Petrusic, W. M. (1994). The calibration of confidence in perceptual judgements. *Perception & Psychophysics, 55*, 412–428

Bar-Hillel, M. (1980). The base rate fallacy in probability judgements. *Acta Psychologica, 44*, 211–233

Bjorkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception & Psychophysics, 54*, 75–81

Blanton, H., Pelham, B. W., DeHart, T., & Carvallo, M. (2001). Overconfidence as dissonance reduction. *Journal of Experimental Social Psychology, 37*, 373–385

Brenner, L. A., Koheler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgements: A critical examination. *Organizational Behavior and Human Decision Processes, 65*, 212–219

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*, 4745–4765

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review, 89*, 306–318

Celebrini, S., & Newsome, W. T. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *Journal of Neuroscience, 14*, 4109–4124

Festinger, L. A. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row & Peterson

Fischoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception & Performance, 3*, 552–564

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology, 52*, 464–473

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435

Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Science, 1*, 78–82

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57*, 226–246

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswickian origins of uncertainty in judgement: A sampling model of confidence in sensory discrimination. *Psychological Review, 104*, 344–366

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review, 103*, 582–591

Keren, G. (1988). On the ability of monitoring non-veridical perceptions and uncertain knowledge: Some calibration studies. *Acta Psychologica, 77*, 217–273

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79*, 216–247

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 107–118

Kunda, Z. (1990). The case for motivated inference. *Psychological Bulletin, 108*, 480–498

Mahajan, J. (1992). The overconfidence effect in marketing management predictions. *Journal of Marketing Research, 29*, 329–342

McGraw, A. P., Mellers, B. A., & Ritov, I. (2004). The affective costs of overconfidence. *Journal of Behavioral Decision Making, 17*, 281–295

Miceli, M., & Castelfranchi, C. (2002). The mind and the future: The (negative) power of expectations. *Theory & Psychology, 12*, 335–366

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*, 143–148

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193–210

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*, 806–820

Part IV
Experimental Interactive Decision Making

Granny Versus Game Theorist: Ambiguity in Experimental Games

J. Eichberger(✉), D. Kelsey, and B.C. Schipper

Keywords: Knightian uncertainty · Choquet expected utility · Equilibrium under ambiguity · Strategic uncertainty · Experiments

1 Introduction

In standard game theory, strategic uncertainty in games is resolved in Nash equilibrium, at least for games with a unique Nash equilibrium. Given a player's equilibrium conjecture about opponents' play, she chooses a best response that conforms to the opponents' equilibrium conjecture about her play. What if players lack confidence in their equilibrium conjectures about opponents' play? This is plausible especially if the game is one-shot and players lack previous experience with the same opponents. Lack of confidence in probability judgements is modelled formally by the literature on ambiguity or Knightian uncertainty (Schmeidler, 1989; Gilboa & Schmeidler, 1989; Bewley, 1986). Recently, such approaches have been applied to strategic games (Dow & da Costa Werlang, 1994; Eichberger & Kelsey, 2000; Marinacci, 2000).[1] Results on the comparative statics of equilibrium under ambiguity have been derived that should at least in principle be testable (Eichberger & Kelsey, 2002, 2005; Eichberger, Kelsey, & Schipper, 2006).

To our knowledge, we present a first attempt to analyze *strategic ambiguity* experimentally. We design an experiment with two-player games, in which we try to introduce ambiguity by varying the identity of the subjects' opponent. Depending on the treatment, subjects have to make choices against a granny, a game theorist or

J. Eichberger
Wirtschaftstheorie I, Alfred-Weber-Institut für Wirtschaftswissenschaften, Grabengasse 14, 69117 Heidelberg, Germany
e-mail: juergen.eichberger@awi.uni-heidelberg.de

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.
© Springer-Verlag Berlin Heidelberg 2008

against some fellow subjects. We find more ambiguity averse behavior when subjects face the grandmother compared to the game theorist. However, there does not seem to be a significant difference between behavior against other subjects and behavior against the grandmother.

The main goal of the experiment is a test of results on the comparative statics of equilibrium with respect to changes in ambiguity. In games with strategic complements and positive externalities, equilibrium actions decrease when there is more ambiguity. The same holds for games with strategic substitutes and negative externalities. The intuition is straight forward. As example of the latter class of games, consider a two-person bargaining game. Players face ambiguity over the share of the pie which the opponent will claim. An ambiguity-averse player puts a high weight on bad outcomes, i.e., the event that the opponent demands a large share. As a result, her best-response is to claim a low share. If ambiguity increases, the best-response demand decreases.

In experiments, it is difficult to control for a subject's ambiguity. We vary, therefore, cardinal payoffs of a game monotonically keeping the ordinal payoff structure constant. In this way we make games increasingly sensitive to ambiguity-averse behavior. We assume that a decision maker facing ambiguity evaluates an action by the Choquet expect payoff, i.e., she forms expectations with respect to possibly non-additive beliefs. By changing the relative size of cardinal payoffs in a suitable way we can manipulate Choquet expected payoffs such that a given degree of ambiguity has a larger effect on behavior. With this procedure we find that our experimental results are in line with the theoretical predictions for the games we analyze.[2]

The paper is organized as follows: The next section introduces briefly the concept of strategic ambiguity behind our study. Section 3 describes the design of the experiment, followed in Sect. 4 by a formal statement of hypotheses and the experimental results. Appendix 1 contains a translation of the instructions.

2 Ambiguity in Strategic Games

Consider a finite two-player strategic game $\Gamma = \langle (A_i)_{i=1,2}, (u_i)_{i=1,2} \rangle$ where A_i is player i's finite set of actions and $u_i : A_i \times A_{-i} \longrightarrow \mathbb{R}$ is player i's payoff function. Each player's ambiguity over the opponent's choice of actions is interpreted as a lack of confidence in a probability assessment over opponent's actions. We assume that each player is a Choquet expected utility maximizer. More precise, a player's beliefs are represented by a capacity on A_{-i}, i.e., a real-valued function $\nu_i : 2^{A_{-i}} \longrightarrow \mathbb{R}$ that satisfies monotonicity, for $E, F \subseteq A_{-i}$, $E \subseteq F$ implies $\nu_i(E) \leq \nu_i(F)$, and normalization, $\nu_i(\emptyset) = 0$ and $\nu_i(A_{-i}) = 1$.

In order to compute the Choquet expected payoff given a capacity ν_i, we order the payoffs of each action a_i from highest to lowest, $u_i^1(a_i) > ... > u_i^k(a_i) > ... > u_i^K(a_i)$. Moreover, we denote by $A_{-i}^k(a_i) := \{a_{-i} \in A_{-i} : u_i(a_i, a_{-i}) \geq u_i^k(a_i)\}$

the set of actions of the opponent which yield better payoffs than $u_i^k(a_i)$ with the convention $A_{-i}^0 := \emptyset$. Player i's Choquet expected payoff from action a_i given her capacity ν_i is the Choquet integral (for details on Choquet expected utility theory, see Schmeidler (1989)),

$$U_i(a_i, \nu_i) := \sum_{k=1}^{K} u_i^k(a_i)[\nu_i(A_{-i}^k(a_i)) - \nu_i(A_i^{k-1}(a_i))].$$

Let the support of a capacity supp ν_i be defined as in Dow and da Costa Werlang (1994) and Eichberger and Kelsey (2000, 2002). More formally, supp ν_i is defined as the set $E \subseteq A_{-i}$ such that $\nu_i(A_{-i} \setminus E) = 0$ and $\nu_i(F) > 0$ for all F such that $A_{-i} \setminus E \subsetneqq F$. There are several notions of support of a capacity used in the literature.[3] We use this notion here in order to be comparable with the literature cited above.

An *equilibrium under ambiguity* of a finite two-player strategic game Γ is a tuple of capacities $(\nu_i^*)_{i=1,2}$ such that for $i = 1,2$ there exists a non-empty support supp ν_i^* for which

$$\text{supp } \nu_i^* \subseteq \arg\max_{a_{-i} \in A_{-i}} U_{-i}(a_{-i}, \nu_{-i}^*).$$

This definition is due to Dow and da Costa Werlang (1994). In equilibrium under ambiguity, the support of each player's equilibrium capacity is a subset of the opponent's best responses given the opponent's equilibrium capacity. In two-player games, if beliefs are additive, then an equilibrium under ambiguity coincides with a Nash equilibrium.

Capacities can be partially ordered by their ambiguity (see Eichberger and Kelsey (2002) and Marinacci (2000)).[4] A game has *strategic complements* (respectively *strategic substitutes*) if there exists an order on the action sets such that each player's best-responses are increasing (respectively decreasing) in the opponent's action a_{-i} on A_{-i}. A game has *positive* (respectively *negative*) *externalities* if there exists an order on the action sets such that $u_i(a_i, a_{-i})$ is increasing (respectively decreasing) in a_{-i} on A_{-i} for all $a_i \in A_i$ and all players. For games with both properties, we require that those properties use the same order on the action sets. Eichberger and Kelsey (2002, 2005) and Eichberger et al. (2006) have shown the following results on the comparative statics of equilibrium with respect to ambiguity for players who are ambiguity averse.[5] If a game has strategic complements and positive (respectively negative) externalities, then equilibria under ambiguity are decreasing (respectively, increasing) in ambiguity. The same holds for games with strategic substitutes and negative (respectively, positive) externalities. Moreover, in games with strategic complements and multiple equilibria, sufficient ambiguity selects among equilibria. Rather than reproducing these results formally, we will illustrate them by an example of the class of 3×3 games which we also use in the experiment.

Example Consider the class of 3×3 games

	X	Y	Z
A	c,b	c,c	$c,0$
B	$0,b$	e,c	$e,0$
C	a,d	d,c	d,e

with $0 < a < b < c < d < e$. This asymmetric game has a unique pure Nash equilibrium, (B,Y). If we define an order $A < B < C$ and $X < Y < Z$, then it is easy to verify that this asymmetric game has strategic complements and positive externalities. Given a capacity ν, compute the Choquet expected payoffs of the row player for her three actions:

$$U(A,\nu) = c$$
$$U(B,\nu) = e\nu(\{Y,Z\})$$
$$U(C,\nu) = a + (d-a)\nu(\{Y,Z\}).$$

Suppose ν is such that $U(A,\nu) < U(B,\nu)$. Then there exists a more ambiguous capacity ν' with $\nu'(\{Y,Z\}) < \nu(\{Y,Z\})$ such that $U(A,\nu') > U(B,\nu')$. This is the case if and only if $\nu(\{Y,Z\}) > \frac{c}{e} > \nu'(\{Y,Z\}$. So, best-responses are decreasing in ambiguity.

In the experiments we try to manipulate the ambiguity for the same strategic game by letting subjects play against different opponents. There is no theory that tells us how to tie ambiguity to the identity of an opponent. In order to elicit how a given player, faced with the same opponent, responds to more ambiguity, we manipulate the cardinal payoffs of games keeping the ordinal payoff structure fixed such that the choice becomes more sensitive to ambiguity. This can be done by manipulating e and c such that $\frac{c}{e}$ changes relative to $\nu(\{Y,Z\})$. The more ambiguity-averse a subject is, the more likely she will choose A rather than B as the ratio $\frac{c}{e}$ falls.[6] □

In the experiment, subjects face three classes of strategic games: Firstly, games with strategic complements, positive externalities and a unique pure Nash equilibrium, henceforth "*strategic complements*", secondly, games with strategic substitutes, negative externalities and a unique pure Nash equilibrium, henceforth "*strategic substitutes*", and thirdly, games with strategic complements and multiple equilibria, henceforth "*multiple equilibria*". There are four 3×3 versions of each class of games for which cardinal payoffs vary monotonically keeping the ordinal payoff structure constant (see Table 1).

The identification numbers of the games are in the top left corner of each game matrix. Games 1–4 in Table 1 are games with strategic complements and positive externalities if we fix the order $A < B < C$ and $X < Y < Z$. They have a unique pure Nash equilibrium, (B,Y). In these games, A is the equilibrium action under ambiguity if ambiguity is sufficiently high, i.e., $\nu(\{Y,Z\})$ is less than the critical value $\frac{c}{e}$. Notice that the ratio $\frac{c}{e}$ increases from game 1 to game 4. The effect of ambiguity on these games has been discussed in the Example.

Table 1 Experimental games

Strategic complements

1.	X	Y	Z
A	25, 23	25, 25	25, 0
B	0, 23	100, 25	100, 0
C	3, 27	27, 25	27, 100

2.	X	Y	Z
A	71, 69	71, 71	71, 0
B	0, 69	100, 71	100, 0
C	3, 73	73, 71	73, 100

3.	X	Y	Z
A	86, 84	86, 86	86, 0
B	0, 84	100, 86	100, 0
C	3, 88	88, 86	88, 100

4.	X	Y	Z
A	97, 95	97, 97	97, 0
B	0, 96	100, 97	100, 0
C	3, 99	99, 97	99, 100

Strategic substitutes

5.	X	Y	Z
A	3, 3	3, 0	27, 25
B	0, 3	100, 100	100, 25
C	25, 27	25, 100	25, 25

6.	X	Y	Z
A	3, 3	3, 0	72, 70
B	0, 3	100, 100	100, 70
C	70, 72	70, 100	70, 70

7.	X	Y	Z
A	3, 3	3, 0	88, 86
B	0, 3	100, 100	100, 86
C	86, 88	86, 100	86, 86

8.	X	Y	Z
A	3, 3	3, 0	99, 97
B	0, 3	100, 100	100, 97
C	97, 99	97, 100	97, 97

Multiple equilibria

9.	X	Y	Z
A	25, 25	25, 0	25, 0
B	0, 25	23, 25	27, 0
C	0, 25	0, 27	100, 100

10.	X	Y	Z
A	70, 70	70, 0	70, 0
B	0, 70	68, 70	72, 0
C	0, 72	0, 72	100, 100

11.	X	Y	Z
A	86, 86	86, 0	86, 0
B	0, 86	84, 86	88, 0
C	0, 88	0, 88	100, 100

12.	X	Y	Z
A	97, 97	97, 0	97, 0
B	0, 97	95, 97	99, 0
C	0, 97	0, 99	100, 100

Games 5–8 are games with strategic substitutes and negative externalities if we fix the order $A > B > C$ and $X > Y > Z$. They also have a unique pure Nash equilibrium, (B,Y). For high ambiguity, C is the only equilibrium action under ambiguity. The more ambiguity-averse a subject is, the more likely she will choose C in these games. Since the critical value increases from game 5 to 8, we should observe more subjects choosing C in this order of the games.

Finally, Games 9–12 are games with strategic complements, positive externalities and multiple equilibria if we fix the order $A < B < C$ and $X < Y < Z$. The pure-strategy Nash equilibria of these games are (A,X) and (C,Z). For a sufficiently high degree of ambiguity, however, only (A,X) is an equilibrium under ambiguity. Notice that this equilibrium under ambiguity coincides with the Pareto-dominated Nash equilibrium. As one moves from game 9 to 12 the critical value for the choice of the ambiguity-averse action increases.

Table 2 Sensitivity to ambiguity

Choice	Capacity	Games			
		Strategic complements			
		1	2	3	4
A	$v(\{Y,Z\}) <$	0.27	0.72	0.86	0.97
		Strategic substitutes			
		5	6	7	8
C	$v(\{Y,Z\}) <$	0.27	0.70	0.86	0.94*
		Multiple equilibria			
		9	10	11	12
A	$v(\{Y,Z\}) <$	0.27*	0.71*	0.86*	0.97*

Table 2 provides the critical values for which ambiguity changes the equilibrium behavior in the twelve games considered.[7]

3 Design

The experiment was computerized using z-tree[8]. In each treatment, subjects played the twelve games described in the previous section. For each game, they had to make a choice of action and indicate their belief about the opponent's action. They did not receive any feedback about the opponent's choice of action after each game. We distinguish three treatments:

Treatment gt

In this treatment subjects were asked in each game to make choices of an action twice in the row player's position, one against a grandmother and one against a game theorist.[9] Both, the granny and the game theorist, were real people.[10] Their choices were recorded prior to the experiment with a paper-based questionnaire. The subjects knew that. Both, the granny and the game theorist took the column player's position. They knew that they were playing against subjects from subject-pool of the Bonn Laboratory for Experimental Economics (mostly students). Until the very end of the experiment, subjects did not know either the choices of the game theorist or the granny.

In addition to making choices, each subject was asked to state which actions of the respective opponent she did "take into consideration for her choice". The answers to this question provided us with information about the strategies of the opponent, which subjects believed to be relevant for their choice. We take these "stated beliefs" as a proxy for the support of the subjects' beliefs about their opponents' behavior.

Treatment g

In this treatment subjects played only against the grandmother. Hence, they had to make only one choice of action in each game. Otherwise this treatment is identical to Treatment **gt**.

Treatment s

In Treatment **s** subjects were playing against each other. An equal number of subjects was selected for the row player position and the column player position. In each game, each subject made a single choice of action against another subject. Subjects did not know the identity of the opponent. For computing payoffs, players were matched randomly with an opponent. In all other respects this treatment is identical to the Treatments **gt** and **g**.

For our method of testing ambiguity, we need to assume that ambiguity does not change during the experiment. Hence, special efforts were undertaken in order to avoid learning effects. First, we provided no feedback about the opponent's choices between games. Second, we made comparisons between games difficult. We feared that if similarities of ordinal payoffs are recognized, subjects analyze the games only a few times and then "log in" to a particular default action. Subjects could not compare the games by clicking back and forward. They faced the games in a random order.[11] Moreover, the games' payoff structure was disguised by adding a randomly chosen small positive constant to each player's payoff.[12] These constants perturb the cardinal payoffs slightly, they make the games asymmetric but keep the ordinal payoffs constant. In addition, subjects had to solve a payoff-irrelevant memory task between games. For this task, they had to memorize a couple of digits displayed for 5 s and repeat them on the next screen. There is evidence in experimental cognitive psychology (Miller, 1956) that humans' short-term memory span is limited to a few digits only. With this memory task we wanted to erase the short-memory of previous games, thus making comparisons more difficult.

Prior to the experiment, subjects received written instructions in German in which the experimental setting was explained in detail (see Appendix 1 for a translation). According to the instructions, subjects knew that they were to make choices in 12 games against a *granny*, a *game theorist*, or *other subjects*, respectively. Subjects were, however, not informed about the types of games which they were to play. In order to be convincing in our claim that the grandmother and the game theorist were indeed real people we provided subjects with additional information about their background. E.g., we informed subjects that the granny is old, raised eight children, and lives in a village in East-Germany and that the game theorist is a successful professor.

The instructions contained also an example of a game which did not belong to the classes of games which they would face in the experiment. With this example we tested prior to the experiment whether subjects understood how payoffs in a game are derived given the choices of the players. The instructions contained also the exchange rate, for which payoffs where exchanged into EURO at the end of the experiment.

At the end of the experiment, subjects had to fill in a brief questionnaire at the computer. The questionnaire contained questions about profession, gender, prior knowledge of game theory or economics, as well as how ambiguous they felt about opponents' choices. Subjects did not know the questions of the questionnaire when they played the games.

Once all the information was collected, three games were randomly selected, their outcomes were computed, converted into EUROS, and paid to the subjects immediately after the experiment. The same holds for the granny and the game theorist, except that they received their payoffs several days after their choices. The subjects' answers to our questions were not rewarded. The experiment lasted for approximately half an hour and subjects earned on average EUR 10.50.

The participants of our experiment were 54 subjects from the subject-pool of the Bonn Laboratory for Experimental Economics, a grandmother and a game theorist. All, but one, subjects of the Bonn Laboratory reported that they were students. About 36% were students of economics or mathematics. Approximately 24% had participated in a course on game theory. Of the students, 36% were female.

The granny and the game theorist were approached directly by the experimenter. We collected the data from the granny and the game theorist a couple of days prior to the experiment. The students' experiment was conducted in the Bonn Laboratory for Experimental Economics in June 2004.

We had 18 subjects for each treatment. Since choices were not revealed until the very end of the experiment, we have 18 independent observations for each treatment. Because the games are not symmetric, however, only the 9 observations of the row players in Treatment **s** are comparable with the observations from the other treatments.

4 Hypotheses and Results

Ambiguity about the behavior of the opponent as modelled by the Choquet expected utility approach described in Sect. 2 induces predictable behavior in games. Our first set of hypotheses and results concern the question whether there are any measurable effects of ambiguity about the opponent's behavior. Our second set of hypotheses and results deals with these comparative statics predictions.

We know from previous experiments on ambiguity in single person decision problems (Camerer & Weber, 1992) that the majority of subjects behave in an ambiguity-averse manner. Hence, we maintain the assumption that subjects behave ambiguity-averse throughout this experiment.

4.1 Is There Ambiguity?

Our motivation for treatments with a grandmother and a game theorist comes from the fact that behavior of subjects may in general not be ambiguous enough to produce observable effects. A priori it is not clear why the behavior of a grandmother should be more ambiguous than that of a game theorist. Given the subject pool of students at the University of Bonn, however, who in some cases have had some experience with game theory, our presumption was that these students would feel less ambiguous about the behavior of an expert game theorist than about the behavior of the grandmother, obviously an non-specialist opponent. We tried to re-enforce this "non-specialist" feature of the grandmother by explicitly mentioning in the instructions that the granny, in contrast to the game theorist, had difficulties in understanding the experimental set up.

Based on this assumption we expect that subjects felt more ambiguity playing against the granny than playing against the game theorist in Treatment **gt**, the only treatment where such a direct comparison is possible. Our experimental results provide us both with a subject's self-assessed feeling of ambiguity and with her actual choice of action. This design motivates the following two hypotheses.

Firstly, we consider ambiguity associated with the player. We predict that subjects will report more ambiguity when playing against the granny. Secondly, we look at ambiguity about the opponent's choice of action. We predict that the higher ambiguity regarding the granny's choice is reflected in the stated beliefs about the set of the opponent's actions which a subject considers possible. The more actions of the opponent a player takes into account, the more ambiguity she experiences. Hence, beliefs about the grandmother's choice should be more coarse than beliefs about the game theorist's choice.

Hypothesis 1 *In Treatment* ***gt****,*

(i) subjects report more ambiguity about the behavior of the granny than about the behavior of the game theorist,
(ii) stated beliefs about the grandmother's choice of actions are coarser as the stated beliefs about the game theorist's actions.

Secondly, regarding actual behavior, we predict that more subjects choose the more ambiguity-averse action if they face the grandmother. For games with strategic complements (Games 1–4), A is an equilibrium action under ambiguity and, for games with strategic substitutes (Games 5–8), C is an equilibrium action under ambiguity, while B is the unique Nash equilibrium in both cases. In games with multiple equilibria (Games 9–12), actions A and C are Nash equilibrium actions. For high ambiguity, however, only A remains an equilibrium action under ambiguity. We should, therefore, expect that the equilibrium actions under ambiguity, A in Games 1–4 and C in Games 5–8, will be chosen more often against the granny than against the game theorist. Moreover, we would expect to see the unique Nash equilibrium strategy B in Games 1–8 and the Pareto-dominant Nash equilibrium strategy C in Games 9–12 more often played against the game theorist than against the granny.

Hypothesis 2 *In Treatment* **gt**, *we expect to observe the following behavior:*

(i) In games with strategic complements (Games 1–4), more (respectively, less) often action A (respectively, B) is chosen against the granny than against the game theorist.
(ii) In games with strategic substitutes (Games 5–8), more (respectively, less) often action C (respectively, B) is chosen against the granny than against the game theorist.
(iii) In games with multiple equilibria (Games 9–12), more (respectively, less) often action A (respectively, C) is chosen against the granny than against the game theorist.

Turning to our results. In the questionnaire of Treatment **gt** we asked subjects the questions listed in Table 3. These questions relate to the ambiguity associated with the opponent's identity. Table 3 shows that 72% of the subjects feel they can predict the behavior of the game theorist better than that of the grandmother. Consistent with this assessment, 72% of the subjects report that they prefer to play against the game theorist. We can reject the hypothesis that subjects can guess the granny's behavior better than the game theorist's behavior (resp. prefer to play against the granny than against the game theorist) at the 0.05 confidence level using a Binomial test. For the third question, the degree of certainty was measured on an integer scale ranging from 0 to 5, with *complete uncertainty* at 0 and *complete certainty* at 5. Table 3 reveals that, on average, subjects were more certain about the game theorist's behavior with 3.3 than about the grandmother's behavior with 1.6. These averages rather hide the actual extent of the uncertainty, since 10 of 18 subjects were certain or completely certain (4 or 5) about the behavior of the game theorist and just 2 subjects felt uncertain or very uncertain (0 or 1). Even stronger is the rating of the granny, where only one subject was certain or completely certain (4 or 5) about the behavior of the granny and 10 subjects felt uncertain or very uncertain (0 or 1). Using a Wilcoxon Signed Ranks test, we can reject the hypothesis that subjects can guess the behavior of both opponents equally well at the 0.03 confidence level.

To see which of the opponent's strategies subjects considered as important for their choice, we turn to Fig. 1. This figure shows how often subjects reported a non-singleton belief about the opponent's actions. Clearly, stated beliefs differ by opponents. Subjects in Treatment **gt** state more often a non-singleton belief[13] when playing against the grandmother (50%) than when playing against the game theorist (40%). This provides some support for our hypothesis that subjects feel more certain about the behavior of the game theorist. We can reject the hypothesis that subjects

Table 3 Perceived ambiguity

Question	Game theorist	Granny
1. Whose behavior can you guess better?	72%	28%
2. Whom would you prefer to play against?	72%	28%
3. How certain are you about the behavior of...?	3.3	1.6

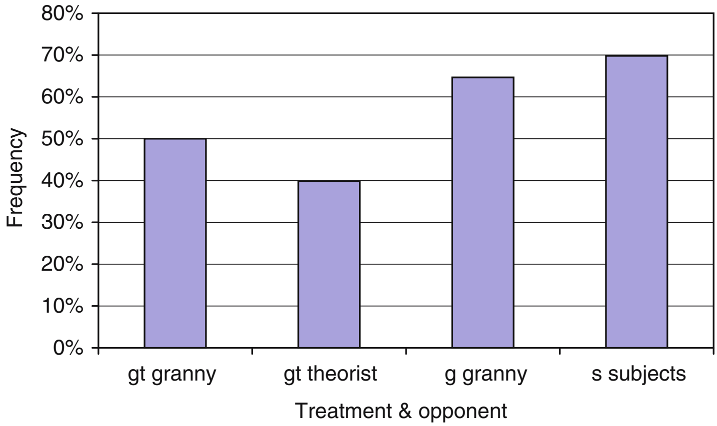


Fig. 1 Non-singleton support of stated beliefs

stated equally often a coarse belief (i.e., two or more actions) for the game theorist and for the granny at a 0.05 confidence level using a Wilcoxon Signed Rank test.

We summarize these results in Observation 1.

Observation 1 *In Treatment* **gt***, we can not reject Hypothesis 1:*

(i) Subjects reported significantly more ambiguity about the behavior of the granny than about the behavior of the game theorist.

(ii) Stated beliefs about the grandmother's choice are significantly more often coarser than stated beliefs about the game theorist's choice.

Figure 2 provides information on Hypothesis 2. In the games with strategic complements, in the upper diagram of Fig. 2, 36% of subjects chose the equilibrium action under ambiguity *A* against the grandmother, while only 21% chose this action against the game theorist. This difference is significant at a 0.11 level using a Wilcoxon Sign Rank test. This observation is consistent with Hypothesis 2 (i). However, also the unique Nash equilibrium action *B* is chosen more often against the grandmother (by 39% of the subjects) than against the game theorist (by 35% of the subjects), which is contrary to Hypothesis 2(i). This difference though is insignificant (0.38 level).[14] It is surprising how often action *C*, which is neither a Nash equilibrium action nor an equilibrium action under ambiguity, was chosen against the game theorist (by 44% of the subjects) in this sequence of games.[15]

In the games with strategic substitutes in the middle diagram of Fig. 2, action *C* is chosen more often against the grandmother (by 28% of subjects) than against the game theorist (by 18% of subjects), which is significant at a 0.10 level. Strategy *C* is the only equilibrium action under sufficiently large ambiguity. The unique Nash equilibrium action *B* is chosen less often against the grandmother (by 68% of subjects) than against the game theorist (by 78% of subjects). This difference is only significant at a 0.12 level. Both observations are consistent with Hypothesis 2(ii).

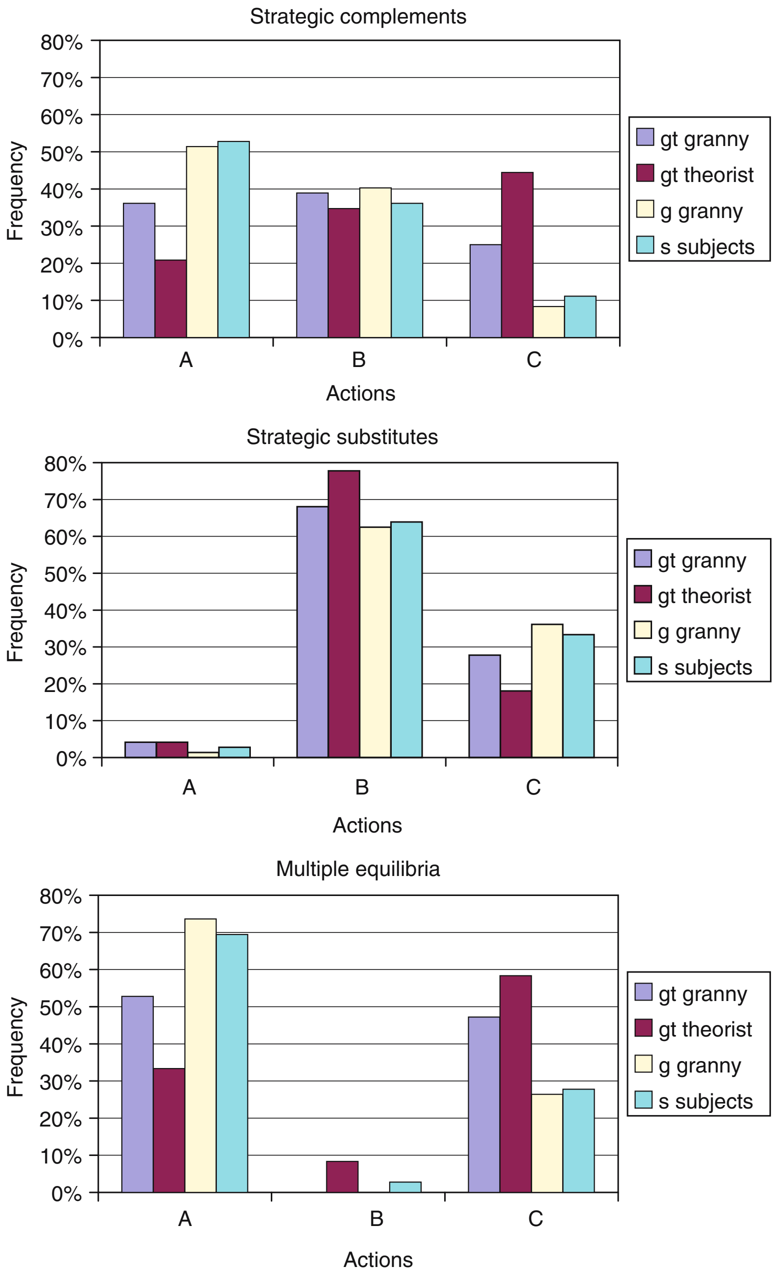


Fig. 2 Distribution of actions

Finally, in the games with multiple equilibria in the lower diagram of Fig. 2, 55% of the subjects chose action *A* against the grandmother, while only 33% of the subjects chose this strategy against the game theorist. This is significant at a 0.08 level. In contrast, action *C* is chosen less often against the grandmother (by 47% of the subjects) than against the game theorist (by 58% of the subjects). This difference is not significant (0.24 level). For low ambiguity, both actions are Nash equilibrium actions and actions in an equilibrium under ambiguity, but if ambiguity is sufficiently large then action *A* becomes the unique equilibrium action under ambiguity. Both observations are consistent with Hypothesis 2(iii).

Observation 2 summarizes these findings.

Observation 2 *In Treatment* **gt***, there is mixed evidence for Hypothesis 2(i), but we cannot reject Hypothesis 2(ii) and Hypothesis 2 (iii):*

*(i) In Games 1–4 (*strategic complements*), subjects chose significantly (resp. insignificantly) more often the ambiguity averse action (resp. the Nash equilibrium action) against the grandmother than against the game theorist.*
*(ii) In Games 5–8 (*strategic substitutes*), subjects chose significantly more (resp. less) often the ambiguity averse (resp. Nash equilibrium) action against the grandmother than against the game theorist.*
*(iii) In Games 9–12 (*multiple equilibria*), subjects chose significantly more (resp. insignificantly less) often the ambiguity averse action (resp. the Pareto-dominant Nash equilibrium) against the grandmother than against the game theorist.*

Treatment **gt** provides the opportunity to compare the behavior of subjects playing games against two opponents with *identical payoffs* but clearly *distinguished by personal characteristics*. The two other treatments, Treatment **g** and Treatment **s**, serve as control treatments. Our a priori hypothesis was that behavior when playing against other subjects should create less ambiguity than playing against the granny, but more ambiguity than playing against the game theorist. Comparing firstly Treatments **g** and **s,** these considerations can be expressed by the following two hypotheses. The first two hypotheses compare stated beliefs and actual behavior in Treatment **g** and in Treatment **s**. Hypothesis 3 is analogous to Hypothesis 1. We assume that the higher ambiguity against the granny compared to the other subjects is reflected in the subjects' statements.

Hypothesis 3 *In comparing Treatment* **g** *and Treatment* **s***, stated beliefs about the grandmother's choice are more coarse compared to stated beliefs about other subjects.*

Hypothesis 4 parallels Hypothesis 2. We predict that the equilibrium action under ambiguity will be chosen more often against the granny than against other subjects, while the Nash equilibrium action will be more often against other subjects.

Hypothesis 4 *In comparing Treatment* **g** *and Treatment* **s***, we expect the following facts:*

(i) In games with strategic complements *(Games 1–4), more (respectively, less) often action A (respectively, B) is chosen against the granny than against other subjects.*
(ii) In games with strategic substitutes *(Games 5–8), more (respectively, less) often action C (respectively, B) is chosen against the granny than against other subjects.*
(iii) In games with multiple equilibria *(Games 9–12), more (respectively, less) often action A (respectively, C) is chosen against the granny than against other subjects.*

The answers to the questionnaire at the end of the experiment hint a first answer to Hypothesis 3. In Treatments **g** and **s**, we asked each subject to rate on a scale from 0 (complete uncertainty) to 5 (complete certainty) how certain he or she was about the behavior of the grandmother or the other subject, respectively. The average reports are very similar in both treatments, 2.6 for Treatment **g** and 2.7 for Treatment **s**.

A similar conclusion can be drawn when looking at stated beliefs in Fig. 1. In fact, there was more ambiguity reported about the choices of the other subject than about those of the granny. However, the difference is not significant (0.27 level using a Wilcoxon–Mann–Whitney test). Observation 3 states this result.

Observation 3 *We can reject Hypothesis 3: Stated beliefs about the grandmother's choice in Treatment* **g** *are less coarse than stated beliefs about other subject's choice in Treatment* **s**.

Examining Fig. 2 shows that also the choices of actions were almost identical in both treatments. In the case of strategic complements (Games 1–4), the equilibrium action under ambiguity was chosen even slightly more often against the other subject than against the granny and the Nash equilibrium action was chosen more often against the granny. The difference between the joint distributions of actions is not significant (0.9 level for strategic complements and substitutes and 0.5 for multiple equilibria using a $\mathcal{X}^2$ test).

Observation 4 *In comparing Treatments* **g** *and* **s**, *we can reject Hypothesis 4:*

(i) In games with strategic complements *(Games 1–4), insignificantly more (resp. less) often action B (resp. A) is chosen against the granny than against other subjects.*
(ii) In games with strategic substitutes *(Games 5–8), insignificantly more (resp. less) often action C (resp. B) is chosen against the granny than against other subjects.*
(iii) In games with multiple equilibria *(Games 9–12), insignificantly more (resp. less) often action A (resp. C) is chosen against the granny than against other subjects.*

Observations 3 and 4 suggest that the perceived ambiguity as well as the actual behavior were similar in Treatment **g** and Treatment **s**. It is important to keep in mind, however, that, for Treatment **s**, this comparison rests on a much smaller number of observations, since only the behavior of the nine row players are used.

Finally, comparing Treatment **gt** and **g**, our ex-ante presumption was that one would find the same perceived ambiguity and the same behavior under ambiguity in regard to the granny in both treatments.

Hypothesis 5 *Choices and stated beliefs when playing against the grandmother in Treatment* ***gt*** *do not differ from Treatment* ***g****.*

In fact, Fig. 1 shows quite clearly that subjects were considering significantly more often non-singleton beliefs when facing the grandmother in Treatment **g** than in Treatment **gt**. This difference is significant at a 0.03 level using a Wilcoxon–Mann–Whitney test. Similarly, Fig. 2 reveals that the ambiguity-related actions, *A* in Games 1–4, *C* in Games 5–8, and *A* in Games 9–12, were chosen more often in Treatment **g** than in Treatment **gt**. The difference between the joint distributions of actions is significant at a 0.05 (resp. 0.02) level for strategic complements (resp. multiple equilibria) but insignificant for strategic substitutes (0.5 level with a $\mathcal{X}^2$ test). To sum up, it appears that subjects felt more ambiguity when playing against the grandmother in Treatment **g** than in Treatment **gt**.

Observation 5 *In comparing Treatments* ***gt*** *and* ***g****, we can reject Hypothesis 5:*

(i) Stated beliefs about the grandmother's choice are significantly more often coarser when playing against the grandmother in Treatment ***g*** *than in Treatment* ***gt****.*
(ii) Play against the granny in Treatment ***gt*** *differed significantly from Treatment* ***g****. In particular, the ambiguity-related actions (resp. Nash equilibrium actions) were more (resp. less) often chosen in Treatment* ***g*** *than in Treatment* ***gt****.*

Observation 5 records stronger ambiguity effects in Treatment **g** than in Treatment **gt**. Though we did expect that playing against the grandmother would create some ambiguity, we were surprised to find this ambiguity to be substantially smaller in the Treatment **gt** where subjects face both the granny and the game theorist. We speculate that this finding is due to a presentation effect. Treatment **gt** is likely to lead subjects towards a comparative judgement between the game theorist and the granny. Such comparative analysis may lead to different judgements of the granny when the granny is the only opponent to judge as in Treatment **g**.

4.2 How Do Subjects React to Ambiguity?

The core hypotheses of this article concern the comparative statics analysis of behavior under ambiguity. As explained in Sect. 2, we constructed the sequence of games in each of the three variants, *strategic complements, strategic substitutes* and *multiple equilibria*, such that the critical level for changing behavior towards the equilibrium action under ambiguity rose with the number of the game. Table 2 contains these critical levels. In each group of games the sensitivity to ambiguity increased with the number of the games. Hence, we advance the following hypothesis.

Hypothesis 6 *For all treatments, we expect to observe following comparative statics:*

(i) In games with strategic complements*, choices of action A (respectively, B) increase (respectively, decrease) from Game 1 to 4.*
(ii) In games with strategic substitutes*, choices of action C (respectively, B) increase (respectively, decrease) from Game 5 to 8.*
(iii) In games with multiple equilibria*, choices of action A (respectively, C) increase (respectively, decrease) from Game 9 to 12.*

Turning now to our results, the left diagrams of Fig. 3 show how the frequency of the equilibrium action under ambiguity changes in all treatments and against all opponents. With the exception of Treatment **gt**, we find for each class of games that the equilibrium action under ambiguity increases as the games become more ambiguity-sensitive, i.e., from the lower to the higher game number. The exception is play against the game theorist in games with strategic complements and against

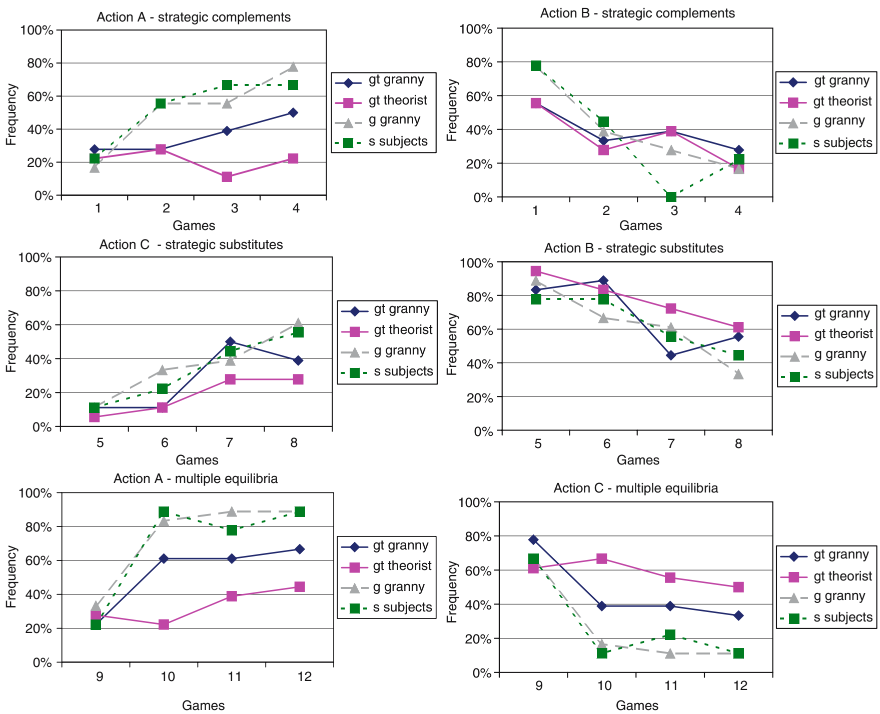


Fig. 3 Comparative statics

the granny in games with strategic substitutes, where in Games 3, 8 and 10 a decline has to be noted.

The right diagrams in Fig. 3 show the frequency of choice for the unique Nash equilibrium action *B* in response to increasingly ambiguity-sensitive games. With the exception of Games 3, 8, and 11 we observe a decrease in all treatments and against all opponents.[16]

We test the results on the comparative statics by comparing each subject's choices in Game 1 versus Game 4 (and similarly Game 5 vs. Game 8 and Game 9 vs. Game 12). We exclude all observations of actions that are neither an equilibrium action under ambiguity nor a Nash equilibrium. We test as the null-hypothesis that switches from the Nash equilibrium to the ambiguity-averse action and vice versa are equally likely. We can reject this hypothesis except for play against the game theorist in games with strategic complements and in games multiple equilibria in Treatment **gt**.[17] Summarizing these results, we obtain Observation 6.

Observation 6 *We can not reject Hypothesis 6, except for play against the game theorist in games with strategic complements and multiple equilibria (Treatment* ***gt****).*

5 Concluding Discussion

In experiments on single-person decision problems, ambiguity plays a role as many studies of the Ellsberg-paradox show. Camerer and Weber (1992) provide a survey of these results. Strategic problems are usually even more complex, so it appears reasonable to assume that ambiguity plays an even bigger role in strategic games. Camerer and Karjalainen (1994) report experiments on strategic versions of Ellsberg's two and three color experiments which seems to confirm this presumption. In their experiments ambiguity concerns the payoffs of the opponents. They find evidence that a substantial fraction of behavior is inconsistent with the assumption of additive beliefs over opponents' types.

To our knowledge, we present a first attempt to analyze *strategic ambiguity* experimentally. By varying the identity of the opponent, we try to introduce different levels of ambiguity in strategic games. Moreover, by varying the cardinal payoffs but keeping the ordinal payoff structure constant, we make games more or less sensitive to the given amount of ambiguity in the experiment. We find that both varying opponents and varying the payoff structure have effects predicted by the theory on ambiguity in games.

In Treatments **gt** and **g**, we used "loaded" instructions in the sense that we described the background of the granny and the game theorist in order to be convincing in our claim that these opponents were indeed real persons. It is therefore justified to ask whether social motives could have brought about the observed difference in choices against the granny and the game theorist. A preliminary check suggests that social motives such as altruism or inequity aversion will induce behavior which is opposed to the one predicted by ambiguity aversion. Thus, they may in fact strengthen our comparative statics results.

Appendix 1: Example of Instructions: Treatment gt (Translation)

Welcome to the Experiment

You participate in an experiment on decision making. You can earn some money. Your earnings depend on your decisions as well as the decisions of a grandmother and a game theorist. Latter decisions we recorded already prior to the experiment today.

The Grandmother

The grandmother is 84 years old. She lives beside a forest in a village in Saxony. She comes from a farmer's family and raised eight children. She likes to take care of her large garden, to solve crossword puzzles, and to watch TV. She faced some difficulties with understanding today's experiment.

The Game Theorist

Game theory is a mathematical theory of strategic decision making such as today's experiment on decision making. The game theorist is Professor of Economic Theory at the University of Bonn. Previously, he worked at Stanford University and Humboldt University, Berlin. He earned a diploma in mathematics and a Ph.D. in economic theory. He published quite a number of articles on game theory in international journals such as the Journal of Economic Theory. He didn't face any difficulties with understanding today's experiment.

Your Decision

Your goal is to maximize your earnings through your choice. You will face decision problems like in the following example (Fig. 4):

You have three actions (A, B, and C), which are marked as rows in above table. The other participant (the grandmother or the game theorist) has three actions as well (X, Y, and Z) (the columns in above table). The numbers in the cells of the table indicate the possible payoffs, whereby your payoff is always the first number in front of the semicolon (;) of each cell, whereas the second payoff belongs to the other participant. For example, if you choose A and the other participant chooses Y, then you receive 56 Taler and the other participant 99 Taler.[18]

Under the table to the left you are supposed to choose your action: One action against the grandmother and one against the game theorist. Prior to your decision, we naturally do not inform you, how the grandmother and the game theorist chose against you. Your payoff depends as indicated in above table on your choice and the choice by the grandmother and the game theorist.

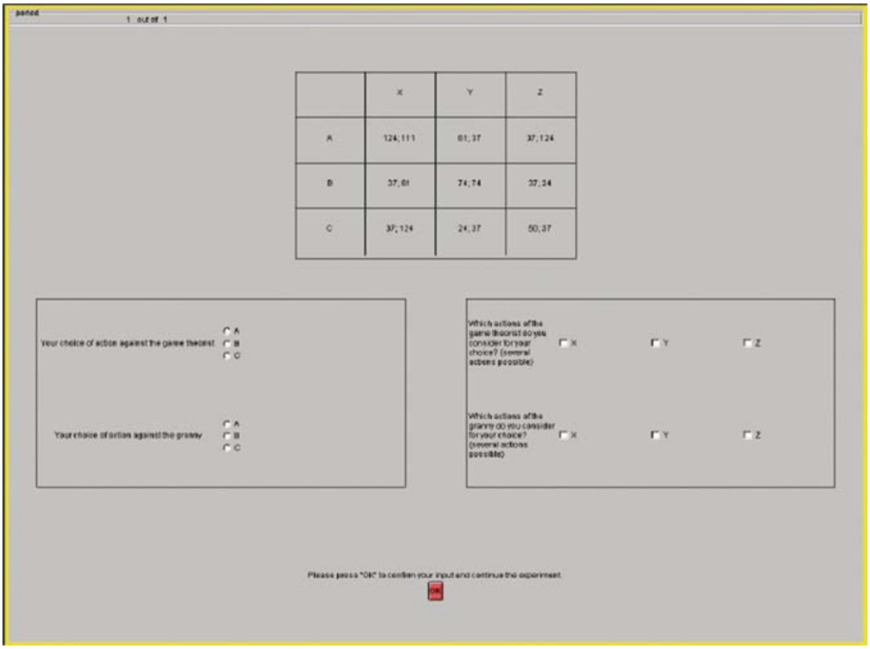


Fig. 4 Screen-shot

Under the table to the right you are supposed to mark the actions that you can not rule out for the grandmother and the game theorist. These are the actions for which you assume that they could be eventually chosen by the grandmother or the game theorist. Here it is possible to mark several actions.

After you made your selection, click "O.K.", and the experiment is continued with a memory task on a new screen. The memory task does not influence your payoff but serves just as an intermediate step between the decision making situations. A sequence of numbers is displayed to you for 5 s, which you should try to remember. After 5 s you are asked on a new screen to reproduce the sequence. After the memory task a new screen appears with a new decision making situation analogous to above. In total there are 12 decision making situations.

Your Earnings

After the decision making situations follows a brief questionnaire. Then you will be informed about your total earnings. To calculate your total earnings, three decision making situations are selected randomly. For each of these three decision making situations the payoff depends on your decisions and the decision of the grandmother and the game theorist as described above. Your total earnings is the sum of payoffs from the three decision making situations against the grandmother as well as the three decision making situations against the game theorist. Your total earnings are exchanged with an exchange rate of 40 Taler = 1 EUR. This amount will be paid to you immediately after the experiment in cash.

In each cabin is a exercise-sheet, which should be completed before the experiment, and which will collected by the experimenter. Only then the experiment will be started. If you have questions now or during the experiment, please quietly contact the experimenter.

Thank you for your participation.

Acknowledgements Financial support by the Deutsche Forschungsgemeinschaft, SFB/TR 15 (Burkhard), the Bonn Graduate School of Economics (Burkhard), by the ESRC grant no. RES-000-22-0650 (David) and by the Deutsche Forschungsgemeinschaft, SFB 504 (Jürgen) is gratefully acknowledged. We thank Robin Cubit, Jörg Oechssler, and seminar participants at the University of Exeter, FUR 2006 and EEA-ESEM 2006 for very helpful discussions.

Notes

[1]See also Epstein (1997); Groes, Jacobson, Sloth, and Tranæ (1998); Haller (2000); Klibanoff (1996); Lo (1996, 1999); Mukerji (1997), and Ryan (2002).

[2]Apart from studying ambiguity, our results may be of independent interest for analyzing experimentally to what extent the opponent's identity has a systematic effect on subjects' play in strategic games.

[3]For different support notions of capacities compare Haller (2000); Marinacci (2000) and Ryan (2002).

[4]Formally, a capacity ν_i'' is reflects more ambiguity than a capacity ν_i' if for all nonempty $E \subsetneqq A_{-i}$, $\nu_i''(E) + \nu_i''(A_{-i} \setminus E) < \nu_i'(E) + \nu_i'(A_{-i} \setminus E)$.

[5]Ambiguity aversion is modelled by the Choquet integral of a *convex capacity*. Formally, a capacity is convex if, for all $E, F \subseteq A_{-i}$, $\nu_i(E) + \nu_i(F) \leq \nu_i(E \cup F) + \nu_i(E \cap F)$.

[6]For the class of games considered above, there is one caveat. We strongly prefer games in which no action is weakly dominated by another (note that ambiguity respects dominance). Thus we also need to increase d whenever we increase c. This influences the evaluation of action C in comparison to action B. However, action A will be preferred to action C.

[7]Numbers with * are just sufficient conditions. For the computations, we took into account the small random constants added (see Note 12).

[8]We are grateful to Urs Fischbacher for making the experimental software available to the profession (Fischbacher, 1999).

[9]For a screen-shot of this treatment see Fig. 4 in Appendix 1.

[10]We want to emphasize that our experiment did not involve any deception of subjects. All the information about the granny and the game theorist provided to the subjects were true.

[11]The games were presented in the following order: 2, 7, 9, 4, 6, 1, 11, 8, 12, 3, 5, 10.

[12] The following table contains the constants which were added.

	Game											
Player	1	2	3	4	5	6	7	8	9	10	11	12
Row	3	2	1	0	3	1	2	1	3	2	1	0
Column	1	1	0	1	1	2	1	3	0	1	2	1

[13]These averages are calculated for all subjects who stated a belief. In 4% of the cases subjects did not state a belief at all.

[14]We have to note a caveat: Since Treatment **gt** concerns a one sample treatment with dependent variables, we could not test for the difference of the joint distributions of *A*s and *B*s.

[15]We suspect that this obviously "irrational" behavior against the game theorist may be a consequence of the random order of games, since the choice of *C* was most pronounced in games following a multi-equilibria game where *C* was the equilibrium action of the Pareto-dominant Nash equilibrium.

[16]Notice that the exceptions seem to occur in the same games.

[17]The significance levels using a Sign Test are given by

Treatment	Strategic complements	Strategic substitutes	Multiple equilibria
gt granny	0.03	0.02	0.01
gt game theorist	0.25	0.11	0.23
g	0.01	0.01	0.01
s	0.03	0.12	0.03

[18]Note that Fig. 4 contains a translated screen-shot in which numbers do not correspond to the translation of the instruction. This is not the case in the German original.

References

Bewley, T. F. (1986). *Knightian decision theory: Part I* (Cowles Foundation Discussion Paper 807)

Camerer, C., & Karjalainen, R. (1994). Ambiguity-aversion and non-additive beliefs in non-cooperative games: Experimental evidence. In B. Munier & M. Machina (Eds.), *Models and experiments in risk and rationality* (pp. 325–358). Dordretch: Kluwer

Camerer, C., & Weber, M. (1992). Recent developments in modelling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty, 5*, 777–795

Dow, J., & da Costa Werlang, S. R. (1994). Nash equilibrium under Knightian uncertainty: Breaking down backward induction. *Journal of Economic Theory, 64*, 305–324

Eichberger, J., & Kelsey, D. (2000). Non-additive beliefs and strategic equilibria. *Games and Economic Behavior, 30*, 183–215

Eichberger, J., & Kelsey, D. (2002). Strategic complements, substitutes and ambiguity: The implications for public goods. *Journal of Economic Theory, 106*, 436–466

Eichberger, J., & Kelsey, D. (2005). Optimism and pessimism in games. (Mimeo, University of Exeter)

Eichberger, J., Kelsey, D., & Schipper, B. C. (2006). Ambiguity and social interaction. Oxford Economic Papers, forthcoming.

Epstein, L. (1997). Preference, rationalizability and equilibrium. *Journal of Economic Theory, 73*, 1–29

Fischbacher, U. (1999). z-Tree – Zurich toolbox for readymade economic experiments – experimenter's manual. (Mimeo, University of Zurich)

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics, 18*, 141–153

Groes, E., Jacobson, H. J., Sloth, B., & Tranæs, T. (1998). Nash equilibrium with lower probabilities. *Theory and Decision, 44*, 37–66

Haller, H. (2000). Non-additive beliefs in solvable games. *Theory and Decision, 49*, 313–338

Klibanoff, P. (1996). Uncertainty, decision, and normal form games. (Mimeo, Northwestern University)

Lo, K. C. (1996). Equilibrium in beliefs under uncertainty. *Journal of Economic Theory, 71*, 443–484

Lo, K. C. (1999). Nash equilibrium without mutual knowledge of rationality. *Economic Theory, 14*, 621–633

Marinacci, M. (2000). Ambiguous games. *Games and Economic Behavior, 31*, 191–219

Miller, G. (1956). The magical number seven, plus or minus two. *Psychological Review, 63*, 81–97

Mukerji, S. (1997). A theory of play for games in strategic form when rationality is not common knowledge. (Mimeo)

Ryan, M. (2002). What do uncertainty-averse decision-makers believe? *Economic Theory, 20*, 47–65

Schmeidler, D. (1989). Subjective probability and expected utility with additivity. *Econometrica, 57*, 571–587

Guessing Games and People Behaviours: What Can We Learn?

A. Morone and P. Morone(✉)

Keywords: Guessing game · p-Beauty contest · Individual behaviour

1 Introduction

In the last decade a growing effort has been devoted to explore the p-beauty contest game (Camerer, Ho and Weigelt, 1998; Duffy and Nagel, 1997; Nagel, 1995; Weber, 2003). The game itself is well known and extremely simple: players are asked to choose a number from a closed interval. The winning player will be the one that gets closer to a target number G. Such target is defined as the average of all guesses plus a constant, multiplied by a real number known to all players. Formally, we can write the target as: $G = p\left(\frac{1}{n}\sum_{i=1}^{n} g_i + d\right)$. In its simplest form the game parameterisation is set as follows: $0 \leq p < 1$, n is the number of players in the contest, $g_i \in [0,100] \subset \mathbf{R}$ is subject i's guess and d is a constant set equal to 0.

Under such definition of G the game-theoretical solution is a unique Nash equilibrium where all players choose 0.[1] In fact, playing 0 is the only strategy that survives the procedure of *iterated elimination of dominated strategies* (IEDS). Let us assume that in the first iteration all players play the highest possible number (100 in our case); here we can immediately observe that the winning number will be $g = p100$. Now, a rational agent should know this and hence play $p100$. However, if all players are rational, the target will shift to $g = p\ (p100)$ or to $g = p^2 100$. Hence, rational players will now play $p^2 100$. This process goes on until the only possible

P. Morone
Department of Economics, Mathematics and Statistics; University of Foggia, Largo Papa Giovanni Paolo II, 1 – 71100, Foggia
e-mail: p.morone@gmail.com

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*.
Theory and Decision Library C 42.

© Springer-Verlag Berlin Heidelberg 2008

equilibrium is reached, i.e. $g = p^{\infty}100 = 0$. Of course, this solution requires that players constantly behave rationally (i.e. for all the infinite iterations of the game) and that everybody knows that everybody else also behaves always rationally. Note that the IEDS suggests what should not be played, and after an infinite number of iterations, the Nash equilibrium is reached.

Nagel, in her seminal paper, suggested "that the 'reference point' or starting point for the reasoning process is 50 and not 100. The process is driven by iterative, naïve best replies rather than by an elimination of dominated strategies" (1995, p. 1325). The *iterative naïve best replies* (INBR) strategy assumes that, at each level, every player believes that he/she is exactly one level of reasoning deeper than all other players.[2] A Level-0 player chooses a number randomly in the given interval [0, 100], with the mean being 50. Therefore, a Level-1 player gives best reply to the belief that everybody else is Level-0 and thus chooses $p50$. Following this line of reasoning, a Level-2 player chooses $p^2 50$, a Level-k player chooses $p^k 50$, and so on. A player who takes infinite steps of reasoning, and believes that all players take (infinite-1) steps, chooses 0, the Nash equilibrium. This interpretation of the converging pattern towards the equilibrium implies that different subjects are characterised by different cognitive levels.

Bosch-Domènech, Montalvo, Nagel and Nagel (2002) analysed 'newspaper and lab beauty-contest experiments' and categorised subjects according to their depth of reasoning. The authors recognised that subjects were actually clustered at Level-1, Level-2, Level-3 and Level-infinity as assumed by Rosemary Nagel.

All these results apply to the standard p-beauty contest game. Under such a standard parameterisation of the game – i.e. $g_i \in [0, 100]$, $p < 1$, and $d = 0$ – both the *iterative naïve best replies* and the *iterated elimination of dominated strategies* require the same number of iterations in order to solve the game. The picture changes if we set $d \neq 0$; in this case the game might well exhibit an interior equilibrium (i.e. different from 0 or 100) and, for specific values of p, the solution of the game obtained, using the two different strategies, involves different numbers of iterations needed to reach the equilibrium.

Güth et al. (2002) proposed a game where d was initially set equal to 0 and subsequently equal to 50. This allowed them to analyse the p-beauty contest from a different perspective, comparing, among other things, interior and boundary equilibria. They showed that the convergence toward the equilibrium is faster when the equilibrium is interior.

In this paper we aim at generalising the *iterative naïve best replies* strategy to the wider class of games with interior equilibria; analyse Güth et al.'s results concerning the properties of interior equilibria in a more general setting; and compare the *iterative naïve best replies* strategy with the *iterative elimination of dominated strategies* for the generalised p-beauty contest. We shall do this by means of a laboratory experiment.

2 A Generalisation of the INBR Strategy to a Game with Interior Equilibria

Let n (>2) be the number of subjects in the game. Each of them has to choose a number $g_i \in [L,H]$, where $L,H \in \mathbf{R}$. Their pay-off function is:

$$u(g_i) = C - c\left|g_i - p\left(\frac{1}{n}\sum_{j=1}^{n} g_j + d\right)\right|,$$

where C is a positive (monetary) endowment, c (>0) is a fine subject i has to pay for every unit of deviation between his/her guess g_i and the target number G.[3] Then, for all $d \in \mathbf{R}$ and $p \in [0,1)$ there is a unique Nash equilibrium, which is given by:

$$\max\left(L,\ \lim_{n\to\infty} Lp^n + \sum_{i=1}^{n} dp^i\right) < g* < \min\left(H,\ \lim_{n\to\infty} Hp^n + \sum_{i=1}^{n} dp^i\right).$$

If we want to solve this generalised p-beauty contest game with the *iterative naïve best replies* strategy we need to redefine it. Since the guessing interval is [L, H], the focal point[4] should be $\frac{H+L}{2}$. The equilibrium using the *iterative naïve best reply* strategy is given by:

$$\begin{cases} if\ p<1 \quad g_i = \max\left(L, \lim\limits_{n\to\infty} \dfrac{H+L}{2}p^n + \sum\limits_{i=1}^{n} dp^i\right) \\ if\ p>1 \quad g_i = \min\left(H, \lim\limits_{n\to\infty} \dfrac{H+L}{2}p^n + \sum\limits_{i=1}^{n} dp^i\right). \end{cases}$$

Now we have defined a generalised theory of naïveté which can be applied both to boundary solutions (as it was originally defined by Nagel, 1995) as well as to interior equilibria; this simple generalisation will help us comparing it with rationality theory.[5]

2.1 Rationality Versus Naivety: Posing Our Research Questions

As already discussed, a p-beauty contest game exhibits a unique boundary Nash equilibrium if the target number is $G = p\left(\frac{1}{n}\sum_{j=1}^{n} g_j + d\right)$ and $d = 0$. Under the assumption of rationality, such converging dynamics takes, theoretically, infinite steps. The situation changes if we consider d values which are greater than 0. In this case the converging equilibrium could be boundary as well as interior.

Güth et al. (2002) carried out an experiment aiming at testing the diverse converging equilibria generated, assigning different values to the parameter d. The authors observed in the lab different converging speeds for different model parameterisations; specifically, they compared two treatments characterised by the following parameters: $p = 1/2$, $g \in [0,100]$, $d = 0$ and $p = 1/2$, $g \in [0,100]$, $d = 50$. Conducting a laboratory experiment, the authors observed that the latter treatment converged towards its Nash equilibrium faster than the former. This result counters the fact that the two treatments had the same degree of complexity. Such apparent contradiction was justified by the authors arguing that the observed difference in converging speeds was due to the fact that in the first case (i.e. $d = 0$) the steady state was a *boundary equilibrium* (i.e. 0), whereas in the second case (i.e. $d = 50$) the system converged towards an *interior equilibrium* (i.e. 50). Hence, they concluded that "interior equilibria trigger more equilibrium-like behaviour than boundary equilibria" (2002, p. 223).

Although it seems appealing, this explanation might be misleading. In fact, dropping the assumption of perfect rationality and applying the theory of naïveté generalised in the section above, we can theoretically calculate the converging dynamics and the equilibria obtainable, using Güth et al. (2002) parameterisation and then compare these results to those obtained applying rationality.

We report these results in Table 1 below. Under the assumption of rationality (i.e. *repeated elimination of strictly dominated strategies*), an infinite number of iterations is required independently of the value assigned to d, hence suggesting that the problems have an identical degree of complexity. The picture changes under bounded rationality assumption (i.e. *INBR*): in this case an infinity-order belief is required to reach the Nash equilibrium when $d = 0$, and only a zero-order belief when $d = 50$. In fact, when $d = 50$, subjects immediately play the Nash equilibrium (which in this case is 50) irrespectively to their sophistication level.[6]

This finding suggest that, if we buy Nagel's idea of bounded rationality and apply the generalised theory of naïveté previously derived, Güth et al. (2002) experimental results could be explained by the fact that the two treatments have a different degree of complexity rather than by the intrinsic capacity of triggering equilibrium-like behaviours of interior equilibria.

In short, we are posing here the problem of understanding what the true reason behind the observed difference in converging dynamics is. In what follows we shall attempt to test the robustness of Güth et al. (2002) results by replicating the p-beauty

Table 1 Güth et al. (2002) treatments – *IEDS* versus *INBR*

Step	$p = 1/2, d = 0, L = 0, H = 100$		$p = 1/2, d = 50, L = 0, H = 100$	
	IEDS	INBR	IEDS	INBR
1	$0 < g < 50$	$g = 25$	$25 < g < 75$	$g = 50$
2	$0 < g < 25$	$g = 12.5$	$37.5 < g < 62.5$	$g = 50$
3	$0 < g < 12.5$	$g = 6.25$	$43.74 < g < 56.25$	$g = 50$
4	$0 < g < 6.25$	$g = 3.13$	$46.87 < g < 53.12$	$g = 50$
⋮	⋮	⋮	⋮	⋮
Infinity	$g = 0$	$g = 0$	$g = 50$	$g = 50$

experiment using different parameterisations. Subsequently, we shall focus our attention on Nagel's theory of naïveté, attempting to understand if it holds also for games which display interior equilibria.

3 Aim and Setting of the Experiment

As discussed above, a preliminary target of our experiment is testing the robustness of the hypothesis according to which "[a]n interior equilibrium [...] is supposed to yield smaller deviations of the guesses from the game-theoretic equilibrium than a boundary equilibrium, since participants often try to avoid extreme choices ..." (Güth et al., 2002, pp. 221–222). In order to test for the validity of such hypothesis we will consider a new set of problems' characterisation defined by different parameterisations of the game. Specifically, we shall compare the original parameterisation adopted by Güth et al. with a similar setting where we vary the value of p (set equal to 2/3) and the value of d (set equal to 25 and 50). It is worth noting that, like in the original experimental setting, this new parameterisation produces an interior game-theoretical equilibrium and a boundary one (when the value of d is respectively 25 and 50). If Güth et al.'s result is robust to different model parameterisations, we would expect to observe a faster convergence in the game with interior equilibrium; otherwise, we shall confute the validity of their results for problems' parameterisations different from those originally selected by the authors.

Once addressed this point, we will move on to consider Nagel's theory of naïveté in the case of games with interior equilibria. In doing so, we will study the first-period choices in two games characterised as above (i.e. $p = 2/3$ and $d = 25$ or 50) and in a new game parameterisation where we will vary the interval $[L,H]$, assigning different values to the upper and lower bound. This will allow us to verify the occurrence of Nagel's naïveté also in games with interior equilibria.

3.1 The Design of the Experiment

In each treatment of the experiment there are $n = 32$ subjects divided into eight groups, each of four subjects. In each group subjects have to guess a number in the real interval $[L,H]$. The closer their guess is to the target the higher is the pay-off. As discussed above, the general form of the pay-off function is: $u(g_i) = C - c\left|g_i - p\left(\frac{1}{n}\sum_{j=1}^{n} g_j + d\right)\right|$.

The experiments were run in October 2005 at the Max Planck Institute of Jena. The software of the computerised experiment was developed in z-Tree (Fischbacher, 1998). Jena University students who participated at the experiment were recruited using the ORSEE software (Greiner, 2004). The age of players ranged from 21 to 31 years, the average pay-off paid amounted to 11.95 €

(*sd* = 1.76), the duration of each treatment was 40 min. Groups were formed randomly at the beginning of the experiment and were kept invariant over the whole experiment (i.e. 40 periods).

4 Results and Interpretation

4.1 Studying Converging Dynamics

In this section we will analyse the results obtained in our experiments. However, before moving to present our findings we shall recall results obtained by Güth et al. (2002) which will serve as a reference point to our study. Schematically we summarise Güth et al. results in Table 2.

As we can see, the authors presented two comparable cases and showed how the treatment where the game-theoretical equilibrium is interior, displayed a higher speed of convergence. We shall now present our results and compare them to those obtained by Güth et al.

In Figs. 1 and 2 we report the average guesses in each group for our first and second treatments. These results appear to confute the findings of Güth et al. (2002),

Table 2 Güth et al. (2002) summary of results

	Parameterisation	Game-theoretical equilibrium	Speed of convergence
Güth et al. Treatment 1	$p = 1/2$, $d = 0$, $L = 0$, $H = 100$	Convergence toward a boundary equilibrium ($g = 0$)	Slower
Güth et al. Treatment 2	$p = 1/2$, $d = 50$, $L = 0$, $H = 100$	Convergence toward an interior equilibrium ($g = 50$)	Faster

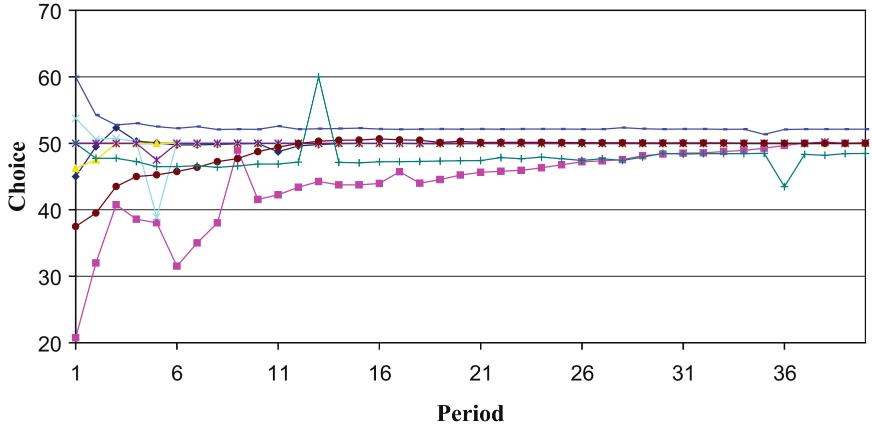


Fig. 1 Treatment 1 – group averages ($p = 2/3$; $d = 25$)

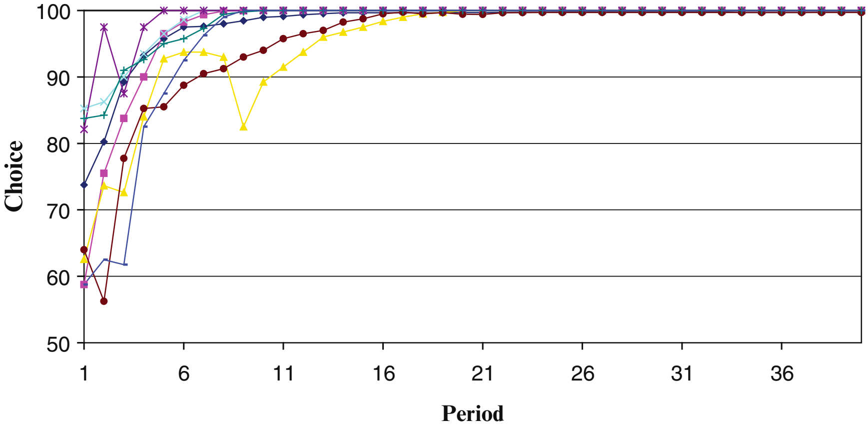


Fig. 2 Treatment 2 – group averages ($p = 2/3$; $d = 50$)

as guesses converge steadily towards the equilibrium, while the interior equilibria treatment converges in a slower pace.

In fact, in treatment 2 the system reaches a steady boundary equilibrium in less than 25 iterations. This result is consistent for each of the eight groups considered in our experiment, a fact which gives a certain degree of robustness to it. On the contrary, not all groups considered in treatment 1 reach a steady equilibrium within the time frame considered (i.e. 40 periods). Moreover, the converging dynamic towards the interior equilibrium is on average statistically significantly slower; more precisely we tested the hypothesis that the two distributions have the same variance. We used the Freund–Ansari–Bradley test. In periods 1, 4, and 6 we rejected the null hypothesis at a significance level of 10% and in periods 2 and 3 we rejected the null hypothesis at a significance level of 5%, in favour of the alternative hypothesis that the variance in treatment 2 is smaller. Hence we can infer that convergence towards equilibrium is faster in treatment 2.

Note that this finding confutes also the generalised version of Nagel's theory of naïveté as also in this case an infinity-order belief is required to reach the equilibrium when $d = 50$, and only a zero-order belief is required when $d = 25$. As is shown in Table 3, following the generalised naïveté theory when $d = 25$, subjects should immediately play the Nash equilibrium irrespectively to their sophistication level. However, this does not happen in the lab.

All in all, this first set of results would suggest a rejection of both Güth et al. (2002) account of convergence (i.e. that interior equilibria trigger more equilibrium-like behaviour than boundary equilibria) as well as our generalisation of Nagel (1995) theory of naïveté, as they proved to be not robust to our new model parameterisation.

This could also imply that Nagel's game-theoretical result cannot be generalised to interior equilibria as it holds solely for boundary equilibria. In what follows, we shall concentrate our attention on first-period choices in order to investigate whether this last hypothesis is actually confirmed by different model parameterisations.

Table 3 Treatments 1 and 2 – *IEDS* versus *INBR*

Step	Treatment 1 $p = 2/3$, $d = 25$, $L = 0$, $H = 100$		Treatment 2 $p = 2/3$, $d = 50$, $L = 0$, $H = 100$	
	IEDS	INBR	IEDS	INBR
1	$16.67 < g < 83.33$	$g = 50$	$33.33 < g < 100$	$g = 66.67$
2	$27.78 < g < 72.22$	$g = 50$	$55.56 < g < 100$	$g = 77.78$
3	$35.18 < g < 64.81$	$g = 50$	$70.37 < g < 100$	$g = 85.19$
4	$43.41 < g < 56.58$	$g = 50$	$80.24 < g < 100$	$g = 90.12$
⋮	⋮	⋮	⋮	⋮
Infinity	$g = 50$	$g = 50$	$g = 100$	$g = 100$

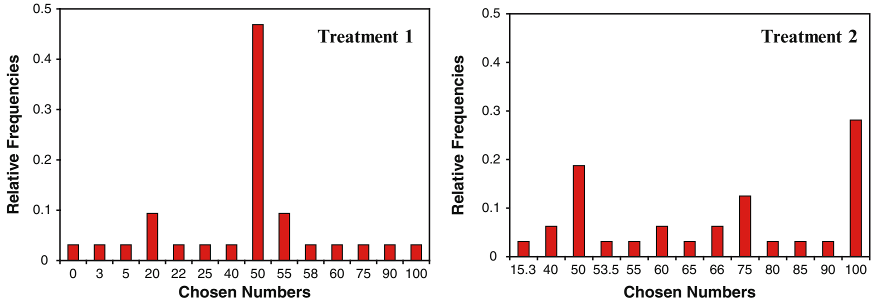
4.2 Studying First-Period Choices

Figure 3 displays first choices frequencies for both treatment 1 and treatment 2 described in the section above. As we can immediately observe, in Fig. 3a almost half of the subjects immediately played the (interior) Nash equilibrium. This might suggest that agents are behaving rationally, as they are instantly able to solve the game applying the *iterated elimination of dominated strategies*. However, we should note that in this very specific case the Nash equilibrium coincides with the salient number calculated following Nagel's definition of a player strategic of degree 0. Moreover, as showed in Table 3, it coincides also with the choice of a person strategic of degree $n \in \mathbf{N}$ (i.e. invariantly of the sophistication level, a person playing the *iterative naïve best replies* always chooses 50).

This implies that, by simply looking at this data, we cannot distinguish among subjects playing 50 as they could be rationally applying the IEDS strategy or they could be as well behaving naively and following an INBR strategy.

We now turn to look at the second treatment. In this case the Nash equilibrium was boundary and equal to 100 and was played in the first period by almost 30% of the players. Note that a person playing strategic of degree 1 would play 66.67; a person strategic of degree 2 would play 77.78, and so on (as reported in the last column of Table 3 above). Not many subjects played strategic of degree 1, 2, 3, ..., as can be easily detected in Fig. 3b. However, almost 30% of them might have played strategic of degree infinite or, alternatively, might have rationally applied the IEDS strategy. Interestingly, almost 20% of the subjects (i.e. 6 out of 32) played 50, which in this case was not a focal point in the sense of being the expected choice of a player who chooses randomly from a symmetric distribution, but was probably perceived as a salient number being the mean of the interval [0, 100]. This fact leads us to believe that when the focal point à la Nagel coincides with a salient number (like the mean of the interval) we might observe players guessing that number for reasons other than playing strategic of degree 0, as suggested by Nagel (1995).

In other words, we shall maintain that Nagel's results might be affected by the specific parameterisation of the model. In order to test this hypothesis we ran two



3a: ($p = 2/3$; $d = 25$; [0, 100]) 3b: ($p = 2/3$; $d = 50$; [0, 100])

Fig. 3 Choices in the first period – Treatments 1 and 2

Table 4 Treatments 3 and 4 – *IEDS* versus *INBR*

Step	Treatment 3 $p = 2/3$, $d = 25$, $L = 13$, $H = 129$		Treatment 4 $p = 2/3$, $d = 50$, $L = 13$, $H = 129$	
	IEDS	INBR	IEDS	INBR
1	$25.33 < g < 102.67$	$g = 64.00$	$42 < g < 119.33$	$g = 80.67$
2	$33.56 < g < 85.11$	$g = 59.33$	$61.33 < g < 112.89$	$g = 87.11$
3	$39.03 < g < 73.41$	$g = 56.22$	$74.22 < g < 108.60$	$g = 91.41$
4	$42.69 < g < 65.60$	$g = 54.15$	$82.81 < g < 105.73$	$g = 94.27$
⋮	⋮	⋮	⋮	⋮
Infinity	$g = 50$	$g = 50$	$g = 100$	$g = 100$

new treatments where the interval boundaries were shifted to the right and were selected as odd integers. Specifically we selected the following interval: [13, 129]. All other parameters were left unaltered. The Nash equilibrium and its converging dynamic are reported in Table 4 for both strategies.

Note that treatments 1 and 3 and treatments 2 and 4 share the same parameters (p and d) and converge to the same Nash equilibrium (in treatment 4, though, the Nash equilibrium is interior whereas in treatment 2 it is boundary). We tested the hypothesis that (T1and T3) and (T2 and T4) come from the same distribution using the Wilcoxon Signed Ranks Test; we can reject the null hypothesis respectively at the 0.0001 and at the 0.001 level. This result suggests that many players do not choose numbers at random but instead are influenced by the value of the boundaries L and H of the game.

In treatments 3 and 4 the game-theoretical Nash equilibrium is always interior and requires an infinity-order belief to be reached independently of the strategy adopted. Looking at Figs. 4a and b we can easily observe that guesses are much less clustered if compared to treatments 1 and 2; further, the number of subjects playing immediately the Nash equilibrium is lower than that observed in Figs. 1 and 2. Specifically, in treatment 3 only 15% of subjects played immediately Nash, and in treatment 4 this share raised slightly to 22%.

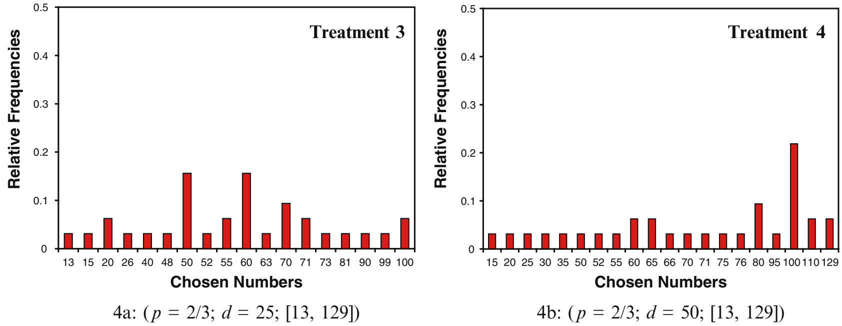


Fig. 4 Choices in the first period – Treatments 3 and 4

As in both treatments there is a very low level of clustering around any focal point, it is hard to believe that agents have been following an *iterative naïve best replies* strategy.[7] However, we shall try to verify if data actually clusters around those iteration levels. In order to test this hypothesis we follow the methodology proposed by Nagel (1995); specifically, we define *neighbourhood* [8] of Step i, where $i \in [0, 1, 2, 3, 4]$.

Intervals between two neighbourhood intervals of Step i and Step $i + 1$ are called *interim intervals*. In Fig. 5 we show the relative frequency of each of these *neighbourhoods* and *interim intervals* for the respective treatment. Note that we cannot define *neighbourhoods* for treatment 1 as the *iterative naïve best replies* strategy leads to the Nash equilibrium at each and every iteration step. Hence, all neighbourhoods would overlap around the game-theoretical equilibrium.

Looking at Fig. 5 we can easily observe that there is not much clustering around iteration levels. The relative frequency is never higher than 15.6%, being on average lower than 6%. Not surprisingly, most of the frequencies are clustered in the upper and lower interim interval. This is certainly due to the fact that these are broader intervals; however, confronting these charts with those reported in Figs. 3 and 4, we can maintain that people tend to cluster initially around round numbers which they perceive as *salient* (like 100, 50 or even 60 and 70 when the guessing space was set as [13, 129]). These findings confirm our assumption that subjects tend to play the focal point when it coincides with other salient numbers of the distribution.

5 Conclusions

In this paper we have addressed the topic of guessing games with the aim of understanding if people play in a rational or naïve way. Two sets of results of the relevant literature triggered our interest. First, Nagel showed how in the first period players deviate strongly from game-theoretic solution. Hence, she proposed a "theory of boundedly rational behaviour in which the 'depth of reasoning' are of importance"

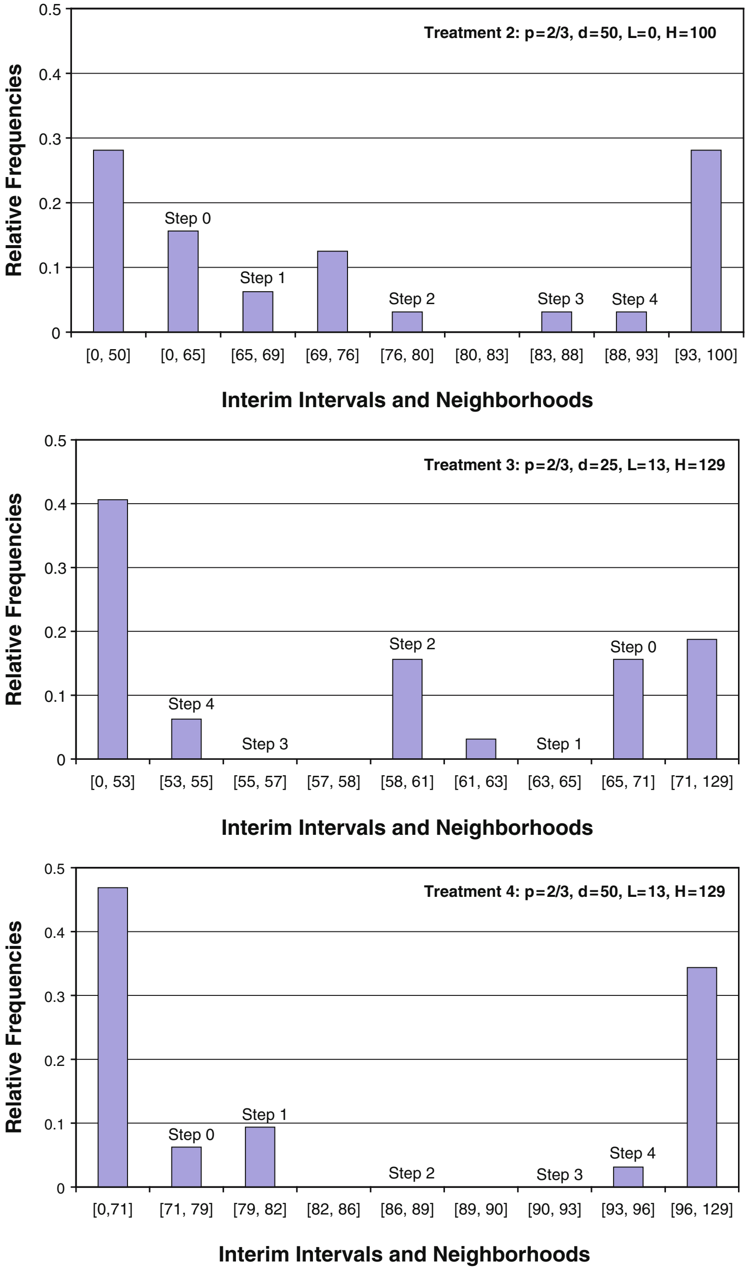


Fig. 5 Relative frequencies of choices in the first period according to the interval classification with reference point *(H + L)/2 + d*

(1995, p. 1325). The author showed that starting from a reference point X (where X was set equal to 50) iteration steps 1 and 2 play a significant role, that is, most of the observations are in the corresponding neighbourhoods. Second, Güth et al. (2002) studied people's behaviour in four different types of experimental beauty-contests. Under the assumption of rational behaviour they found faster convergence to the equilibrium when the equilibrium was interior.

By developing a generalised theory of naïveté (which accounted for interior equilibria) we showed how Güth et al.'s result was compatible with Nagel's theory of boundedly rational behaviour. However, we also wanted to test the sensitivity of both results to different model parameterisations. By conducting a new series of experiments we countered both results showing how, under new parameters, neither the convergence towards interior equilibria was always faster, nor starting from a reference point X (which in our case was different from 50), iteration steps 1 and 2 played any significant role.

We conclude that the results of Güth et al. (2002) and Nagel (1995), however interesting, are severely affected by the ad hoc parameterisation chosen for the game. Far from providing conclusive evidence on the issue of guessing games and people behaviours, this paper aims at raising questions: what are the true driving forces behind subjects decision in a p-beauty contest game? Further, do subject in the lab behave rationally or do they follow naïve strategies? Can we really define a unifying theory of behaviour applicable to all subjects?

Acknowledgements The authors are grateful to Werner Güth and Matthias Sutter for the helpful discussion provided in different stages of the work, and all participants to the FUR 2006. Authors would like to thanks the Max Planck Institute of Jena for funding the experiment reported in this paper. The usual disclaimers apply.

Notes

[1]Note that "[f]or p=1 and more than two players, the game is a coordination game, and there are infinitely many equilibrium points in which all players chose the same number". For $p>1$ and $2p<$ n "all choosing 0 and all choosing 100 are the only equilibrium points. Note that for $p>1$ there are no dominated strategies" (Nagel, 1995, p. 1314).

[2]Please note that in what follows we shall use Level, Step and Degree interchangeably.

[3]Note that this pay-off function was first used in Güth et al. (2002). We prefer it over the standard "winner takes it all" pay-off function as it prevents subjects' income polarisation.

[4]We are borrowing this term from Nagel (1995) where the focal point was ad-hoc set equal to 50.

[5]In what follows, when talking of 'theory of naïveté', we refer explicitly to Nagel's *iterative naïve best replies* strategy and to the *iterative elimination of dominated* when talking of 'rationality theory'.

[6]This can be easily proved numerically. Note that in this very specific case the Nash equilibrium coincides with the "expected choice of a player who chooses randomly from a symmetric distribution" as well as to "a salient number à la Shelling" (Nagel, 1995, p. 1315).

[7]That is, taking (H + L)/2 + d as an initial reference point and considering several iteration steps from this point (Step 0 $\rightarrow$ p [(H + L)/2 + d]; Step 1 $\rightarrow$ p (Step 0 + d); ...; Step i $\rightarrow$ p(Step i-1 + d)).

[8]In general the neighbourhood interval of Step i has the boundaries (Step $i)p^h$ and (Step $i)p^{-h}$, where h is the smallest integer such that two neighbourhood intervals do not overlap. Following Nagel (1995), we rounded intervals upper and lower boundaries to the nearest integers, since mostly integers were observed.

References

Bosch-Domènech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, two, (three), infinity,...: Newspaper and lab beauty-contest experiments. *The American Economic Review, 92*(5), 1687–1701

Camerer, C., Ho, T. H., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental “p-beauty contest”. *The American Economic Review, 88*(4), 947–969

Duffy, J., & Nagel, R. (1997). On the robustness of behaviour in experimental “beauty contest” games. *The Economic Journal, 107*(445), 1684–1700

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178

Greiner, B. (2004). *The online recruitment system orsee 2.0 – A guide for the organization of experiments in economics*. (Working Paper Series in Economics 10). German: University of Cologne, Department of Economics

Güth, W., Kocher, M., & Sutter, M. (2002). Experimental ‘beauty contest’ with homogeneous and heterogeneous players and with interior and boundary equilibria. *Economic Letters, 74*, 219–228

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review, 85*(5), 1313–1326

Weber, R. A. (2003). Learning with no feedback in a competitive guessing game. *Games and Economic Behavior, 44*, 134–144

The Determinants of Individual Behaviour in Network Formation: Some Experimental Evidence

D. Di Cagno(✉) and E. Sciubba

Keywords: Network formation· Experiments· Social interaction

1 Introduction

The role of social networks in shaping economic outcomes has received increasing attention in recent years. Network externalities have been extensively studied both in industrial organisations and, more recently, within the theory of social capital and development economics. Most of this literature however takes the structure of the social network as given and analyses the consequences of network externalities on outcomes.[1]

In this paper we take the view that social linkages are often voluntarily formed and hence the architecture and membership of social networks are part of the economic outcome that one aims to explain. The theoretical literature on endogenous network formation stems from two seminal contributions by Bala and Goyal (2000) and Jackson and Wolinsky (1996). Both papers follow a game-theoretic approach to the formation of social ties where the main idea is that players earn benefits from being connected both directly and indirectly to other players and bear costs for maintaining direct links.

The process of forming a network is extremely complex. The main difference between a network and a series of bilateral links lies in the value that accrues to agents though indirect connections: any two economic agents who have to decide whether to establish a social tie take into account not only their own characteristics and the characteristics of the prospective partner, but also their (and the prospective partner's) position in the social network.

D. Di Cagno
LUISS Guido Carli, Viale Pola 12, 00198 Roma, Italy
e-mail: ddicagno@luiss.it

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*.
Theory and Decision Library C 42.

© Springer-Verlag Berlin Heidelberg 2008

Given that the process of network formation is so complex, predicted outcomes are typically not unique. Even for those cases where the stable network architecture is unique (for example, the star network in information communication models à la Bala and Goyal or Jackson and Wolinsky), the coordination problem of which agent occupies which position in the network still remains.

There is a lively experimental literature on this topic which has mainly focussed on the issue of convergence to a stable network architecture as predicted by the theory. In presence of multiplicity of equilibria and coordination problems, it is hardly surprising that most experimental contributions on this topic have highlighted the difficulty in obtaining convergence. More in detail, while convergence may be more easily achieved in settings where the stable network architecture is the wheel (for positive results see Callander and Plott (2005) and Falk and Kosfeld (2003); for a negative result see Bernasconi and Galizzi (2005)), convergence is always problematic in frameworks where the prediction for the stable network is the center-sponsored star (Berninghaus, Ott, & Vogt, 2004 and Falk & Kosfeld, 2003). Berninghaus et al. (2004) and Falk and Kosfeld (2003) highlight the role of complexity and inequality aversion in preventing convergence to network architectures that are not "fairness compatible" and argue that a network architecture, such as the star network, which results in an uneven distribution of payoffs is less likely to be observed in the lab. Deck and Johnson (2004) avoid coordination failures by introducing heterogeneity among agents and by constructing a framework where the stable network is indeed unique. Vanin (2002) attempts to facilitate coordination by allowing cooperation and by preventing renegotiation among (skilled) subjects: he finds that, even under such favourable conditions, coordination is not achieved in all cases.

Even in absence of coordination, the observed network structures are ultimately the outcome of individual linking decisions. For this reason, in this paper we mainly focus, rather than on convergence, on the possible determinants of individual behaviour.

We run a computerised experiment of network formation, where all connections are beneficial and only direct links are costly. The network formation protocol that we adopt, unlike the one used by most of the experimental literature that has focussed on convergence (see Bernasconi and Galizzi (2005), Berninghaus et al. (2004), Callander and Plott (2005), Falk and Kosfeld (2003), and Goeree, Riedl, & Ule (2005)), requires that links are not unilateral, but have to be mutually agreed in order to form. In particular, players simultaneously submit link proposals and a connection is made only when both players involved agree.[2] Although mainly focusing on individual linking decisions, we provide an analysis both at the macro and the micro level. From a macro perspective, in accordance with the existing literature, we find that convergence to a stable network architecture is made problematic by the presence of multiple equilibria, despite the fact that in our setting links have to be mutually agreed and mutually sponsored. By analysing network formation over time, we nevertheless detect a tendency towards efficient network architectures where all subjects are included and with the minimum number of costly links. At the level of the individual, we estimate the probability of a link through a probit model that includes both best-response and behavioural variables. We find strong evidence that both play a role in network formation. In particular we consider several possible determinants of agents' linking decisions, including best-response behaviour and inertia. We identify two major drives to network formation: optimising best-response behaviour and the attempt to coordinate on the efficient network structure, often through reciprocal and inertial behaviour.

The paper is developed as follows. Section 2 describes the experimental design: the model and the experimental procedure. Section 3 presents the results and Section 4 concludes the paper. The instructions (in their English translation) can be found in Appendix 1. The software[3] utilised for the experiment is available from the authors upon request.

2 The Experimental Design

2.1 The Model

We model network formation as a non-cooperative simultaneous move game. As in Goyal and Joshi (2006) we assume that players' strategies are vectors of intended links and that links are only formed when they are mutually agreed, i.e. desired by both parties involved. There are positive network externalities in that both direct and indirect connections are beneficial; however direct links are costly.

Consider a set N of $n \geq 3$ players, indexed by $i = 1, 2, \ldots, n$. Each player i submits a vector of intended links:

$$s_i = (s_{i1}, s_{i2}, \ldots, s_{in})$$

An intended link is $s_{ij} = 0, 1$ where $s_{ij} = 1$ means that player i intends to link to player j, while $s_{ij} = 0$ means that player i does not intend to link to player j. A link between i and j is formed if and only if $s_{ij} = s_{ji} = 1$. We denote the formed link by $g_{ij} = g_{ji} = 1$, while we represent the fact that there is no mutually agreed link between i and j by setting $g_{ij} = g_{ji} = 0$. A strategy profile for all players

$$s = (s_1, s_2, \ldots, s_n)$$

induces an (undirected) network of links $g = \{g_{ij}\}_{i,j \in N}$, where players are nodes and links are the edges between them. We say that i and j are connected in the graph g if there exists a path of adjoining nodes $k_1, k_2, \ldots, k_m$ such that $g_{ik_1} = g_{k_1k_2} = \ldots = g_{k_{m-1}k_m} = g_{k_mj} = 1$.

Denote by n_i^d the number of direct neighbours of player i, and by n_i the number of his direct and indirect connections. More in detail, denote by n_i^d the number of elements of the set $N_i^d = \{j \mid g_{ij} = 1\}$ and by n_i the number of elements of the set $N_i = \{j \mid \text{there is a path in } g \text{ from } i \text{ to } j\}$. Notice that if i and j are directly linked, then there is a path between them (of length 1): hence necessarily $n_i \geq n_i^d$. Player i's payoff, given his position in the network g, is assumed to be equal to:

$$\pi_i(g) = b \cdot n_i - c \cdot n_i^d$$

where b and c are non-negative constants that represent respectively the unitary benefit from (direct and indirect) connections and the unitary cost of direct links.

Players aim at maximising their payoffs and can rationally form new links or sever existing ones to this aim. Goyal and Joshi (2006) characterise equilibrium networks by introducing the notion of pairwise stable networks. A pairwise stable network is such that there exists a Nash equilibrium strategy profile that induces the network (so that no agent has any incentive to deviate from his current vector of intended links) and such that no pair of agents have any incentive to form a new link. More in detail, for any two agents who are not linked in a pairwise stable network, if one of the two gains by establishing a new link, it must be the case the other player involved is made strictly worse off by the new link. Formally:

Definition: *A network g is a pairwise stable network if the following conditions hold:*

1. *There is a Nash equilibrium strategy profile (s_i^*, s_{-i}^*) that induces g;*
2. *For $g_{ij} = 0$, if $\pi_i(g + g_{ij}) - \pi_i(g) > c$ then $\pi_j(g + g_{ij}) - \pi_j(g) > c$*

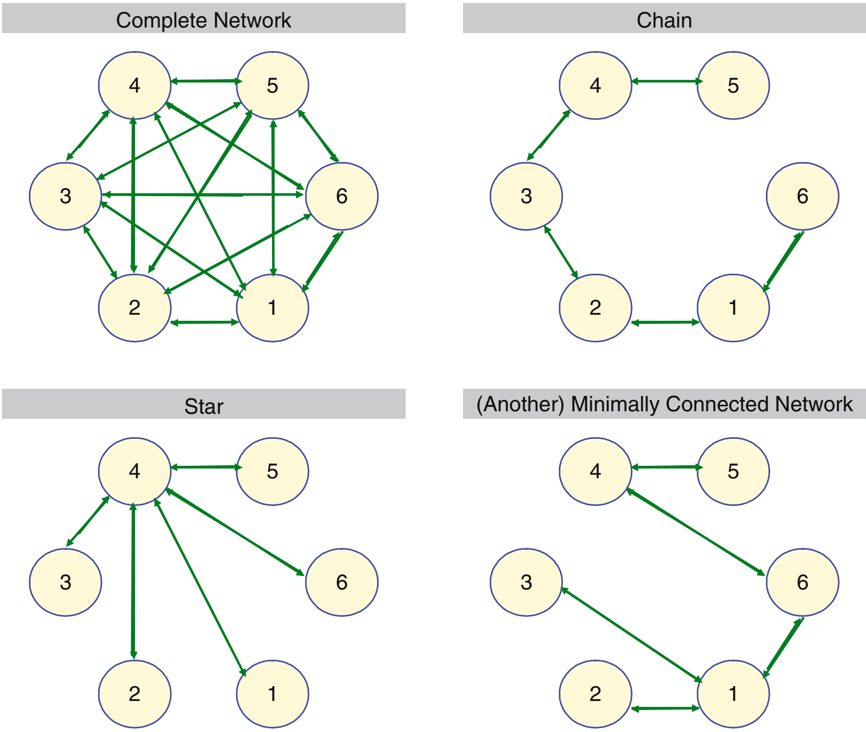
Goyal and Joshi show that all Nash networks are minimal. A minimal graph is such that there is at most one path connecting any two agents: there are no redundant links. The intuition why this has to hold is that if there are redundant links then there are agents that can be reached both directly and indirectly. Players could obtain higher payoffs by deleting their (costly) direct links to all those nodes that they are able to reach indirectly through others.

When $b > c$, then all pairwise stable networks are *both* minimal *and* connected (or minimally connected), i.e. there is one and only one path connecting any two agents. The intuition of why this is so is that if there is any isolated node, given that the benefit from an extra connection is higher than the cost of a direct link ($b > c$), then there are incentives for a new link to be formed between the isolated player and at least another node in the graph.

When $b < c$, the only pairwise stable network (and the only Nash network) is the empty graph.

The complete network, where every node is directly connected to every other, is an example of connected graph. The complete network is clearly not minimal, as there are many redundant links. Examples of minimally connected graphs are the star and the chain.

When the unitary benefit from indirect connections b is lower than the unitary cost of direct links c, but not very much lower, we have a case of conflict between individual incentives to network formation and social optimum. Individual rationality leads to an empty Nash network. However aggregate profits may be higher in a minimally connected network than in the empty graph. More in detail, the aggregate



profits in an empty graph are always zero, while it is easy to check that the aggregate profits in a star network, for example, are strictly positive as long as $b > \frac{2c}{n}$. Given that by assumption $n \geq 3$, this condition can be met for a large enough n even when $b < c$.

2.2 The Experimental Procedure

The experimental sessions were conducted in Spring 2006 at LUISS University in Rome. Subjects were first year Economics students and in total we had 84 participants: 28 women and 56 men, coming from the North (6%), the Center (60%) and the South (34%) of Italy. Each subject participated in only one session and none had previously participated in a similar experiment. We run 14 computerised experimental sessions, with six participants each. Each experimental session lasted between 30 and 45 min. Subjects total earnings were determined by the sum of the profits in each round and were paid using a conversion rate of 100 points per euro. Participant earned approximately 35 € on average.

We implemented two different treatments: sessions 1–7 (treatment 1) involved a lower cost of link formation; sessions 8–14 (treatment 2) involved a higher cost of link formation. Initial endowment and unitary benefit were kept constant across treatments. In more detail, parameters for both treatments are presented in the table below:

	Participants	Initial endowment	Cost	Benefit
Treatment 1 (sessions 1–7)	6	500	90	100
Treatment 2 (sessions 8–14)	6	500	120	100

Under treatment 1 parameters are such that participants cannot run into losses. Under treatment 2 losses were avoided by the fact that the computer would not accept link proposals exceeding the budget constraint. All relevant parameters were equal across participants and displayed on the screen at any time throughout the experiment.

At the beginning of a session subjects were told the rules of conduct and provided with detailed written instructions, which were read aloud by the experimenters. At the end of each session, participants completed a brief questionnaire with basic demographic information.

Sessions consisted of a minimum of 15 rounds, with a random stopping rule determining the end of the experiment.[4] In each round subjects were asked to submit (anonymously and independently) a vector of intended links. The initial screen for each participant is shown in Fig. 1a.

Participants are represented on the screen by different symbols which we considered neutral in that they do not provide subjects any particular clue when deciding to establish a link with another player in the group.[5] Subjects do not know their

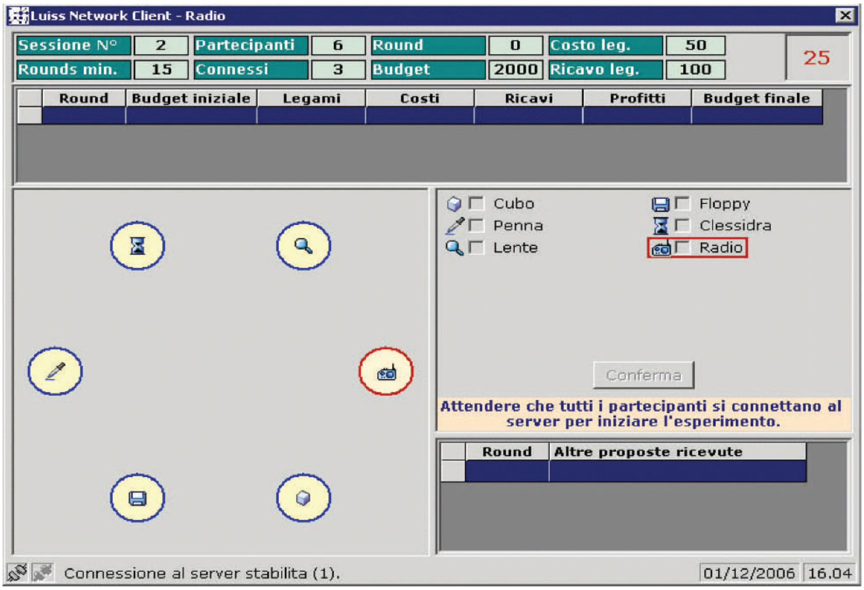


Fig. 1a The initial screen

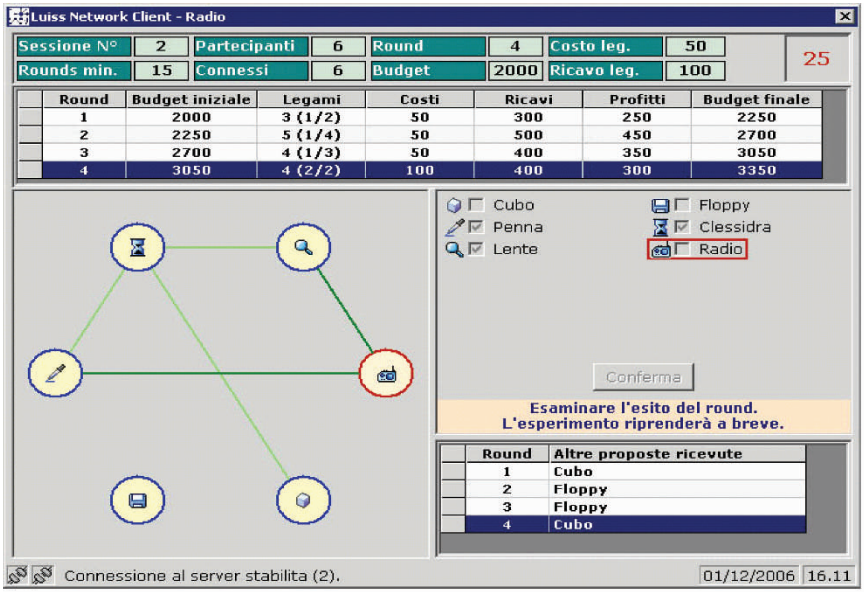


Fig. 1b The participants' screen at the end of round 4

symbol (or the other participants' symbols) in advance and can identify themselves on the screen because their symbol is circled in red. The screen also displays the relevant parameters for the session at play. After all subjects have confirmed their choice of network partners, the computer checks which links are mutually desired and activates them. At the end of each round payoffs are computed and displayed on the screen. Great care was put in making sure that all information available to experimental subjects was provided in a user-friendly way. For this reason the graphical interface was designed so that actual links are visualised on the screen as a graph, rather than as a list of activated ties, or as a matrix of 0/1 connections.

As an example, Fig. 1b shows the participants' screen at the end of round number 4. It displays the graph of all active links, total revenues, costs and profits in the round. It also provides information on unmatched proposals: each subject is informed of those players who have proposed a link to them but whom they have not reciprocated. At any time during the experiment participants have access to a great deal of information on past history: by clicking on the bar corresponding to each round they are able to visualise the graph of active links and the profits obtained in that round.

3 Results

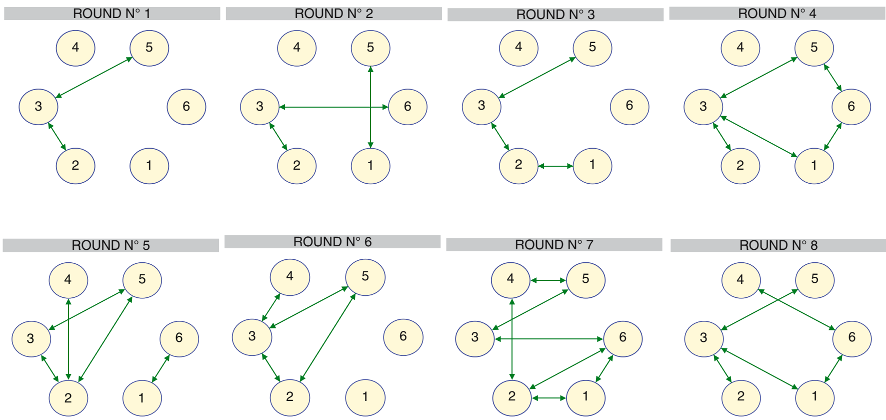
In presenting our results, we distinguish between two different levels of analysis: macro and micro aspects. In the macro analysis we mainly look at the overall resulting network of established links, and at its evolution over time. The number of potential network configurations with six agents is very large: in our macro analysis we focus on whether there is any particular network architecture, among the very many that are possible, that emerges as stable in our experimental sessions and, if it does, on how it compares with the one predicted by the theory. When we move to consider the micro aspects we mainly focus on the analysis of the determinants of individual behaviour for the proposals of new links. In particular we estimate through a probit model the likelihood of link proposals as a function of best response determinants and other behavioural factors.

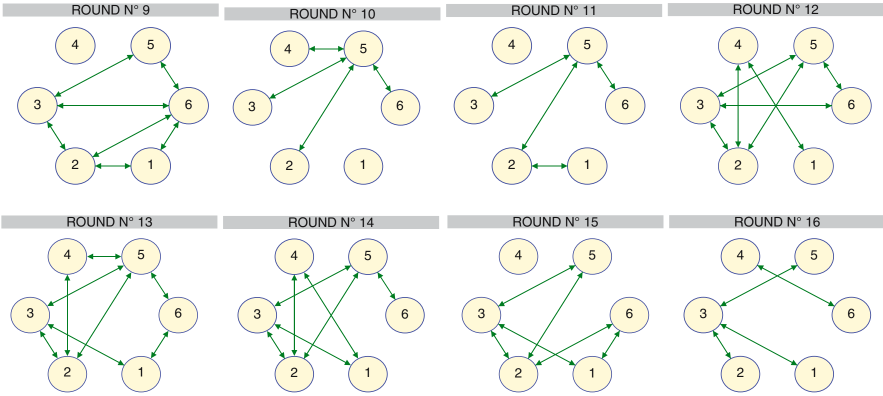
3.1 Macro Aspects

Under treatment 1 any minimally connected graph is a pairwise stable network. Minimally connected graphs are also efficient in that they maximise aggregate profits. Under treatment 2 the only Nash network (and pairwise stable network) is the empty graph. However any minimally connected graph is efficient.

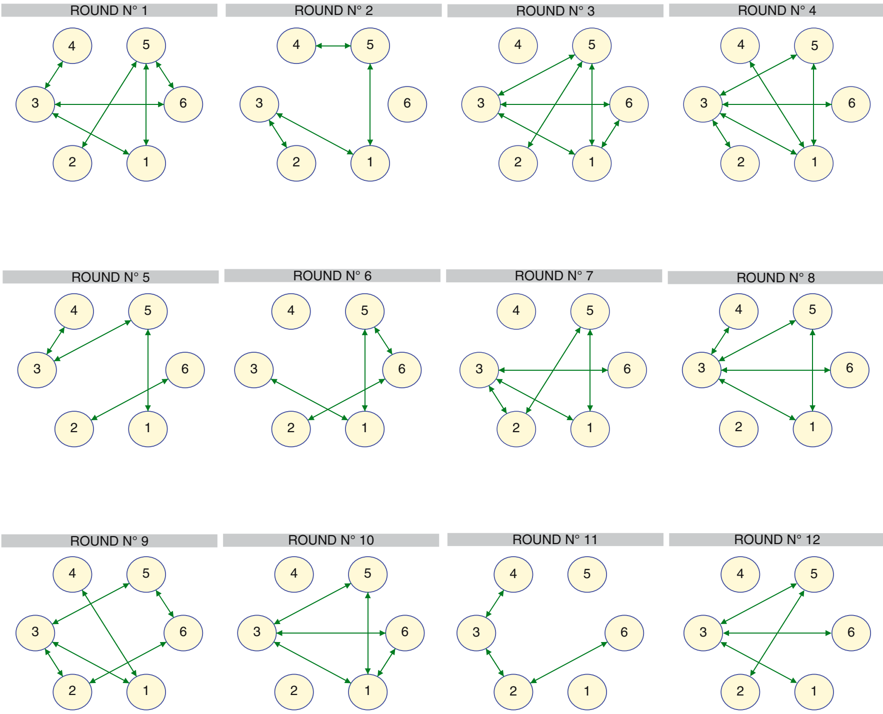
As an example of treatment 1 (sessions 1–7), we show below the macro outcome of experimental session 5. In all sessions under treatment 1 we do not find a definite convergence to a particular network architecture, however we notice that there is a tendency towards connectedness and inclusion. After the first few rounds, it is unlikely to observe isolated nodes. For example, in session 5, for 5 out of 16 rounds (rounds 7, 8, 12, 13, 14) the network formed was connected, with no agent excluded from the network of links. Moreover, there is a tendency to eliminate redundant links: for 7 out of 16 rounds the network was minimal (rounds 1, 2, 3, 8, 10, 11, 16), i.e. any two agents who are connected are reached through a single path. Finally in round 8 the resulting graph is minimally connected and hence it corresponds to a pairwise stable network as predicted by the theory.

As an example of treatment 2, we show below the network formation process for session 9. As far as minimality is concerned, under treatment 2 the outcome

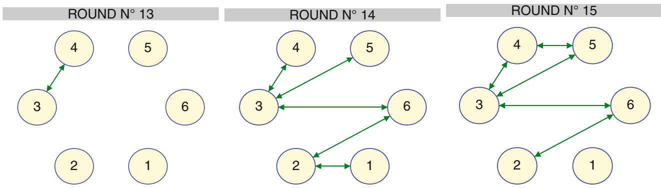




is marginally better than under treatment 1: direct links are expensive, and this is reflected by the fact that there are fewer redundant connections. As one may expect with these parameters, there is less connectedness and the number of isolated nodes is higher than under treatment 1.



We believe that convergence to a minimally connected graph is made very difficult by two main factors. First of all, as it has been remarked by the literature (see



Kosfeld (2004)), the game that agents play has multiple equilibria and players find it very difficult to coordinate on the same Nash equilibrium (clearly communication was prevented during the experiment). Secondly, subjects display some aversion to inertia and, whenever a minimally connected graph is reached in early rounds, it is often abandoned later (in same cases to be reached again) by subjects who cannot resist to the experimentation of new strategies.

Table 1 shows the average number of direct links, indirect connections and failed links by round in each of the two treatments. On average experimental subjects have established 1.64 direct links and 1.52 indirect connections per round in sessions 1–7 and 1.41 direct links and 1.44 indirect connections per round in sessions 8–14.

We also compute the ratio between indirect connections and direct links. Such a ratio captures an important feature of the network architecture: the larger its value, the larger the number of agents that the experimental subject has managed to reach indirectly. The smallest value for this ratio is 0 (no other agent is accessed indirectly: the subject has to bear the entire cost of his/her connections); the largest value for this ratio is 4 (the subject is connected to five others, through a single direct link).

Table 1 Links and connections by round

	Sessions 1 - 7					Sessions 8 - 14				
	Direct	Indirect	Failed	Accpt/Prop	Ind/Dir	Direct	Indirect	Failed	Accpt/Prop	Ind/Dir
1	0.43	0.29	0.93	0.32	0.67	1.29	1.48	0.86	0.60	1.15
2	0.71	0.24	1.17	0.38	0.33	1.24	1.38	1.02	0.55	1.12
3	1.29	1.10	1.07	0.55	0.85	1.43	1.05	1.10	0.57	0.73
4	1.24	1.10	1.14	0.52	0.88	1.43	1.62	0.95	0.60	1.13
5	1.19	0.90	1.24	0.49	0.76	1.38	1.00	1.12	0.55	0.72
6	1.62	1.67	1.14	0.59	1.03	1.33	1.62	1.14	0.54	1.21
7	2.00	1.95	0.93	0.68	0.98	1.33	0.81	1.21	0.52	0.61
8	1.57	1.52	1.19	0.57	0.97	1.57	1.81	0.88	0.64	1.15
9	1.76	1.48	1.10	0.62	0.84	1.48	1.52	1.02	0.59	1.03
10	1.67	1.43	1.17	0.59	0.86	1.43	1.19	1.07	0.57	0.83
11	1.95	2.14	0.95	0.67	1.10	1.52	2.05	0.95	0.62	1.34
12	1.86	2.05	1.12	0.62	1.10	1.29	1.24	1.02	0.56	0.96
13	2.05	2.00	0.95	0.68	0.98	1.48	1.52	0.95	0.61	1.03
14	2.14	2.00	0.76	0.74	0.93	1.57	1.86	0.88	0.64	1.18
15	1.95	1.76	1.00	0.66	0.90	0.95	0.67	1.31	0.42	0.70
16	2.06	1.61	0.94	0.69	0.78	1.27	0.80	0.97	0.57	0.63
17	2.11	2.33	0.94	0.69	1.11	1.44	1.11	0.94	0.60	0.77
18	2.22	1.22	0.78	0.74	0.55	1.44	1.78	0.72	0.67	1.23
19	1.33	2.00	1.17	0.53	1.50	1.33	1.33	0.78	0.63	1.00
20						2.00	3.00	0.50	0.80	1.50
avg	1.64	1.52	1.04	0.60	0.90	1.41	1.44	0.97	0.59	1.00

Here we find that, while there are fewer direct links under treatment 1 than under treatment 2, the number of indirect connections that each experimental subject has managed to secure on average through each of his direct links is higher in the high cost treatment than in the low cost treatment. More in detail, each direct neighbour brings 0.9 indirect connections in sessions 1–7 and 1 indirect connection in sessions 8–14. Moreover there are fewer failed links under treatment 2 (0.97 per round on average), than under treatment 1 (1.04 per round on average). Finally the ratio of accepted links to proposed links is similar under both treatments: only approximately 60% of proposals are reciprocated and result in active links.

Figures 2,3 and 4 show the average number of direct links, indirect connections and failed proposals over time. The number of direct links increases over time, and more so for the low cost treatment, where the average number of links goes from a minimum of 0.43 in round 1, to a maximum of 2.22 in round 18. There are not substantial differences across the two cost treatments as far as the evolution over time of both indirect connections and failed proposals is concerned. The number of failed links declines over rounds, suggesting that experimental subjects learn over time which are the nodes that are likely to reciprocate their proposals.

Table 2 shows the actual number of redundant links by session and by round. We find that the average number of redundant links is much lower under treatment 2 (0.60) than under treatment 1 (1.18). The difference between the two treatments becomes even more stark if we look at the average number of redundant links in the last five rounds of each sessions, when the experimental subjects have clearly acquired more experience of the network formation game. Figure 5 shows how the

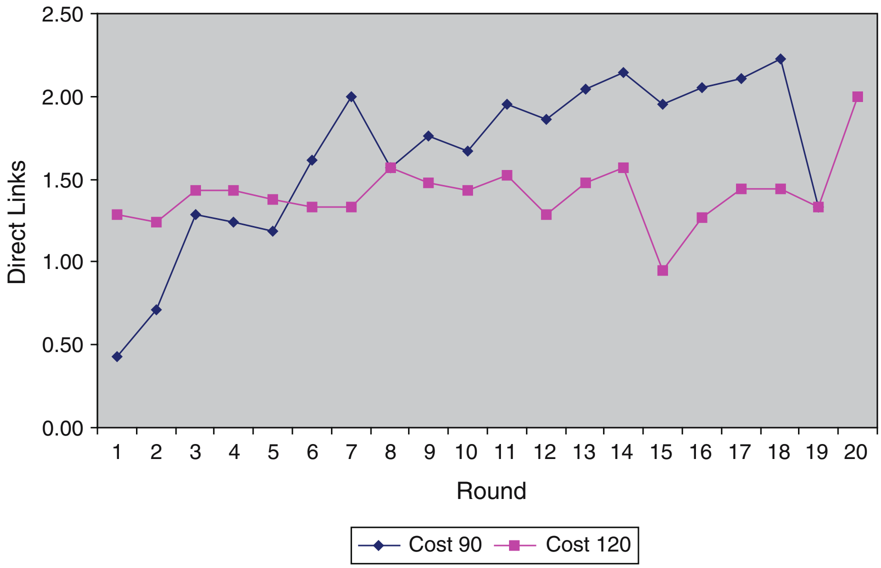


Fig. 2 Average number of direct links by round

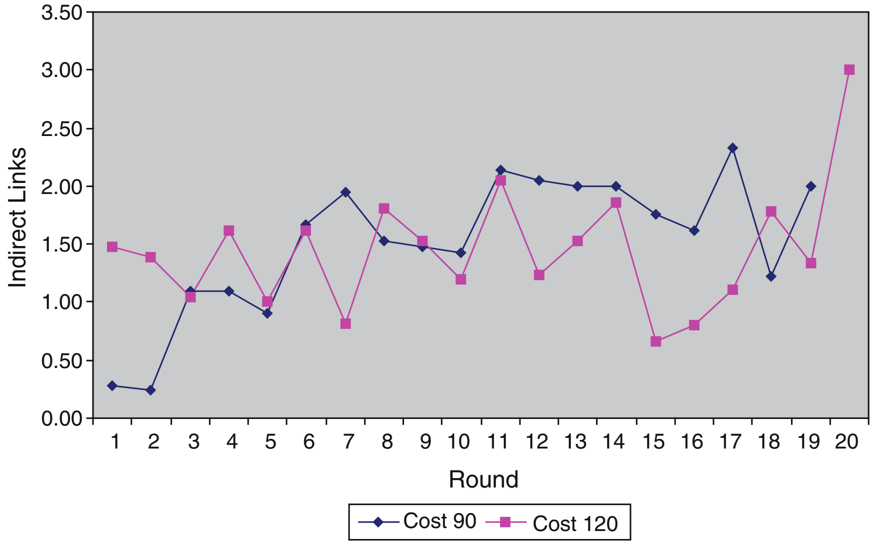


Fig. 3 Average number of indirect links by round

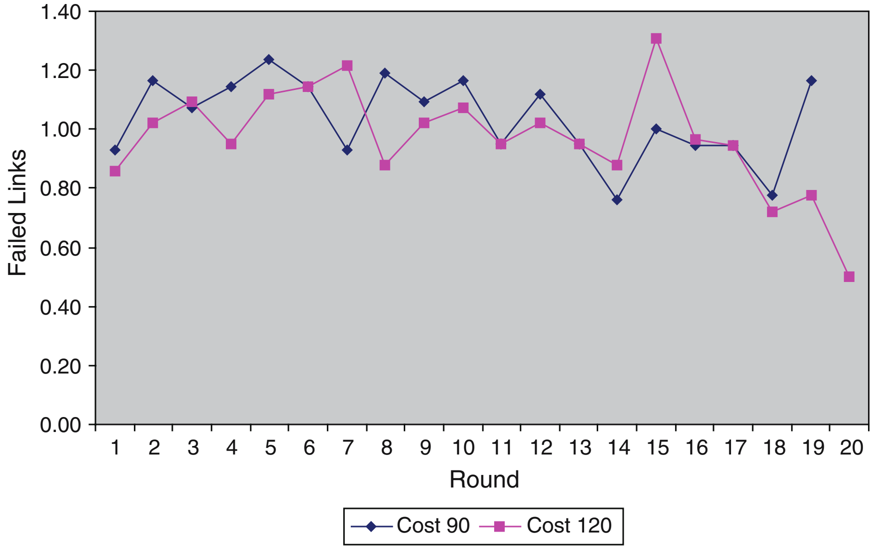


Fig. 4 Average number of failed links by round

number of redundant links evolves over time: rather than decreasing, it increases for sessions 1–7 and it stays stable between values of 0.5 and 1 for sessions 8–14.

Table 3 shows the actual number of isolated nodes by session and by round. While the average number of isolated nodes does not differ greatly across treatments, we do find significant differences when we focus on the average number of

Table 2 Redundant links by session and by round

Redundant Links

	1	2	3	4	5	6	7	avg	8	9	10	11	12	13	14	avg
1	0	0	0	0	0	1	0	0.14	0	2	1	0	0	0	0	0.43
2	0	0	0	1	0	0	0	0.14	0	2	0	0	1	0	0	0.43
3	3	0	0	1	0	0	0	0.57	3	1	1	0	0	0	0	0.71
4	1	0	0	1	1	0	0	0.43	0	1	1	0	1	0	1	0.57
5	1	0	0	0	1	0	1	0.43	0	4	0	0	0	0	1	0.71
6	4	0	0	1	1	1	0	1.00	1	0	0	1	0	1	0	0.43
7	4	0	1	3	2	0	1	1.57	0	1	1	0	1	0	1	0.57
8	4	0	0	3	0	0	1	1.14	1	2	1	0	0	0	0	0.57
9	4	0	0	1	3	1	1	1.43	0	1	1	0	0	0	1	0.43
10	2	3	0	0	0	0	4	1.29	0	2	1	0	0	0	3	0.86
11	2	2	0	3	0	0	3	1.43	2	0	0	1	0	0	0	0.43
12	2	1	0	1	2	0	1	1.00	2	1	0	0	0	1	0	0.57
13	5	3	0	1	3	0	0	1.71	1	2	0	0	0	0	3	0.86
14	4	4	0	4	2	0	0	2.00	0	2	0	0	1	0	1	0.57
15	5	4	0	0	2	0	1	1.71	0	0	1	0	0	0	0	0.14
16	5	4	2	0	0	0		1.83		1		0	0	0	3	0.80
17		4	0			1		1.67				0		0	3	1.00
18		5	2			1		2.67				0		0	1	0.33
19						0		0.00				0		0	2	0.67
20															1	1.00
avg	2.88	1.67	0.28	1.25	1.06	0.26	0.87	1.18	0.67	1.38	0.53	0.11	0.25	0.11	1.05	0.60
last 5	4.2	4.2	0.8	1.2	1.8	0.4	1	1.94	1	1.2	0.2	0	0.2	0	2	0.66

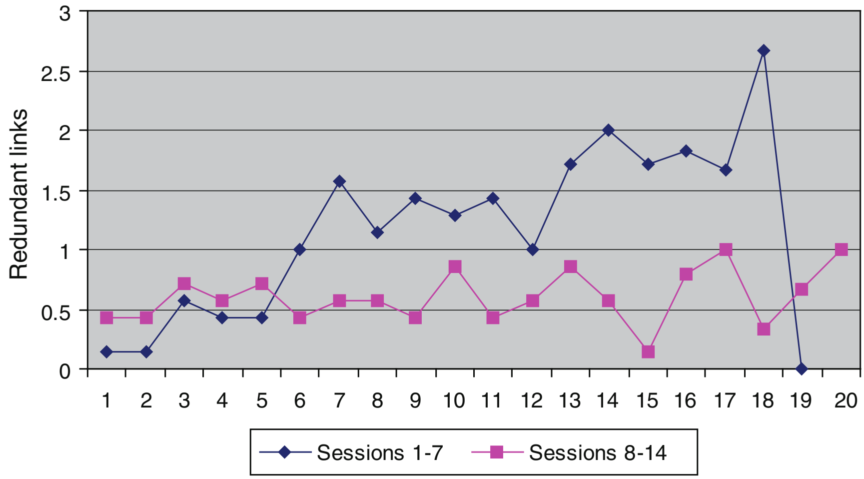


Fig. 5 Average number of redundant links by round

isolated nodes over the last five rounds of each session. Here the low cost treatment results in much more inclusive networks, where only 0.5 nodes are isolated on average. In the high cost treatment the number of isolated nodes does not decrease much over time and equals an average of 1.37 in the last five round. Figure 6 shows the evolution of the number of isolated nodes over time for both treatments.

Efficient networks in our setting are both connected and minimally connected: efficiency requires no isolated nodes and no redundant links. In the network formation

Table 3 Isolated nodes by session and by round

Isolated Nodes

	1	2	3	4	5	6	7	avg	8	9	10	11	12	13	14	avg
1	6	2	6	6	3	2	6	4.43	0	0	0	2	2	6	1	1.57
2	4	4	3	1	1	2	4	2.71	1	0	1	2	1	6	1	1.71
3	1	3	1	0	2	1	2	1.43	0	1	1	0	1	2	2	1.00
4	0	4	0	1	1	1	2	1.29	2	0	0	1	0	2	3	1.14
5	1	2	1	1	0	4	2	1.57	0	0	0	1	2	3	3	1.29
6	0	1	1	0	2	3	1	1.14	0	1	1	0	3	3	2	1.43
7	0	0	1	0	0	2	0	0.43	1	1	1	0	2	1	1	1.00
8	0	3	1	0	0	4	2	1.43	0	0	1	1	1	1	0	0.57
9	0	2	2	0	1	2	1	1.14	0	0	0	0	0	3	1	0.57
10	0	1	2	0	1	2	0	0.86	1	0	2	4	1	3	1	1.71
11	0	0	2	0	1	1	0	0.57	0	1	2	0	0	1	1	0.71
12	1	0	2	0	0	0	0	0.43	1	1	1	4	2	2	1	1.71
13	0	0	2	0	0	0	1	0.43	0	1	4	1	0	1	1	1.14
14	0	0	2	0	0	2	0	0.57	0	0	0	2	1	1	0	0.57
15	1	0	1	0	1	3	0	0.86	4	0	1	2	2	4	0	1.86
16	0	0	1	0	0	1		0.33		2		2	3	2	1	2.00
17		0	0			1		0.33				1		3	1	1.67
18		0	2			1		1.00				0		1	1	0.67
19						1		1.00				2		3	0	1.67
20															0	0.00
avg	0.88	1.22	1.67	0.56	0.81	1.74	1.40	1.18	0.67	0.50	1.00	1.32	1.31	2.53	1.05	1.20
last 5	0.4	0	1.2	0.2	0.2	1.4	0.2	0.51	1	0.8	1.6	1.4	1.6	2.6	0.6	1.37

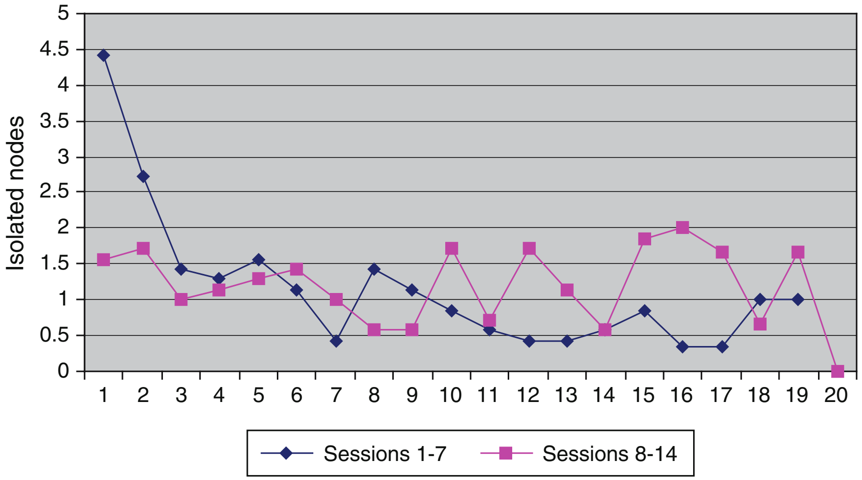


Fig. 6 Average number of isolated nodes by round

process there is clearly a trade-off between these two objectives: when experimental subjects propose many links it is more likely that there are fewer isolated nodes, but it is also more likely that some of the established links are redundant because they link directly and costly to nodes to which they are already indirectly connected. On the other hand, when the network is more sparse and there are fewer direct connections, there are possibly fewer redundant links, while the number of isolated nodes is likely to be higher. A network is more efficient when it obtains higher connectivity

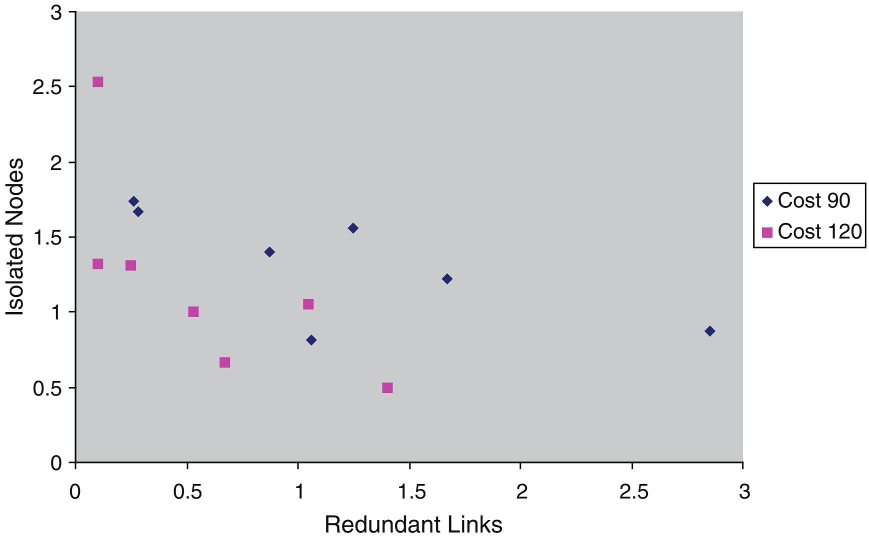


Fig. 7 Trade-off between redundant links and isolated nodes

(i.e. it minimises the number of isolated nodes) by the smallest number of redundant links. Figure 7 shows this trade-off for each of the sessions. We find that the network formation process leads to more efficient networks in the high cost treatment: in sessions 8–14 (unitary cost of direct links equal to 120) the experimental subjects managed to form networks with a smaller number of redundant links and a smaller number of isolated nodes, than in sessions 1–7 (unitary cost of direct links equal to 90).

Table 4 shows the average total profit made by experimental subjects in the two treatments. It is not surprising to find that the average profit in sessions 1–7 (171.93) is higher than in sessions 8–14 (115.96). We also compare the actual profits obtained in the experimental sessions to what would have been obtained by the six participants on average if they had formed minimally connected networks such as the star or the chain. In sessions 1–7, if participants had linked up as a star the average per round profit would have been equal to:

$$\frac{(500-450)+5(500-90)}{6}=350$$

where $(500-450)=50$ is the profit of the hub, and $(500-90)=410$ is the profit of each of the five spokes.[6] Similarly, in sessions 8–14 the average payoff for a star network is equal to:

$$\frac{(500-600)+5(500-120)}{6}=300$$

Table 4 Average profits by round

	Sessions 1 – 7		Sessions 8 –14	
	Profit	% Profit wrt chain or star	Profit	% Profit wrt chain or star
1	32.86	0.09	121.90	0.41
2	32.86	0.09	113.33	0.38
3	124.29	0.36	76.19	0.25
4	123.81	0.35	133.33	0.44
5	103.33	0.30	72.38	0.24
6	185.71	0.53	135.24	0.45
7	219.05	0.63	54.29	0.18
8	171.90	0.49	149.52	0.50
9	168.10	0.48	122.86	0.41
10	161.43	0.46	90.48	0.30
11	236.67	0.68	174.29	0.58
12	225.24	0.64	98.10	0.33
13	223.33	0.64	122.86	0.41
14	225.24	0.64	154.29	0.51
15	198.57	0.57	47.62	0.16
16	222.00	0.63	54.67	0.18
17	254.44	0.73	82.22	0.27
18	144.44	0.41	148.89	0.50
19	213.33	0.61	106.67	0.36
20			260.00	0.87
avg	171.93	0.49	115.96	0.39

We find that while in the low cost treatment participants on average managed to secure profits equal to 50% of the profits obtainable in a minimally connected network, in the high cost treatment profits were equal to less than 40% of those that participants could have achieved if they had for example formed a chain. The networks that were formed in the high cost treatment were less profitable on average, despite being more efficient, in that they presented a superior trade-off of redundant links versus isolated nodes compared to networks formed in the low cost treatment (see Fig. 8). This is mostly due to the lower levels of connectivity obtained in sessions 8–14, i.e. to the larger number of isolated nodes that one observes when the cost of direct links is higher, especially towards the end of each experimental session. Figure 8 shows the evolution of average profits over time: they tend to increase, but at a higher rate for sessions 1–7 than for sessions 8–14.

3.2 Micro Aspects

We move next to a micro analysis of the determinants of individual behaviour in link formation. Such an analysis is particularly valid in this context where because of coordination problems, macro convergence is difficult to observe. In fact, even in

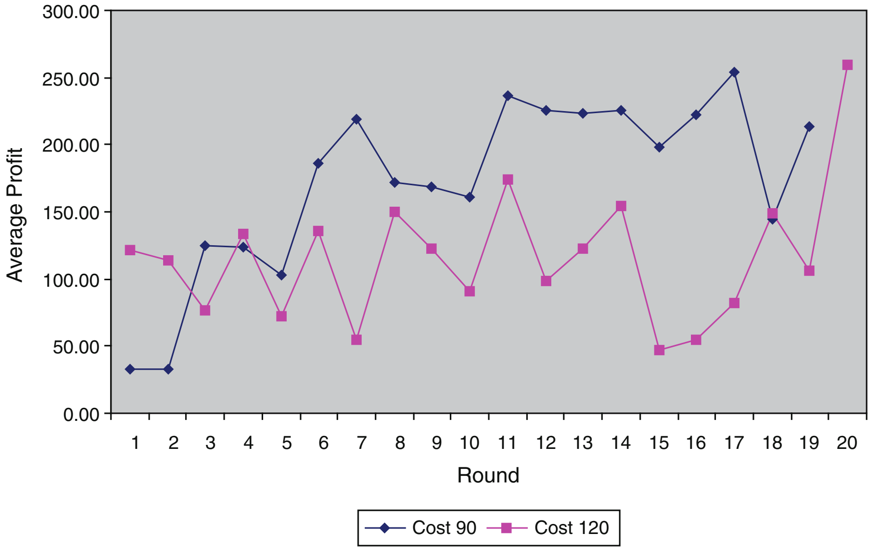


Fig. 8 Average profits by round

presence of mis-coordination, we still ought to be able to determine whether at the individual level subjects are behaving as the theory predicts and what are the main drives to link formation.

Through a probit model we estimate the probability of each subject i proposing a link to each other subject j as a function of the position of i and j in the network of links that were activated in the previous round and that are represented graphically on the subjects' screens. More in detail we estimate the probability of proposing a link as a function of: number of links of the proponent in the previous round; number of links of the recipient in the previous round and cost of link formation. Moreover we include as regressors several binary variables that denote whether the proponent and recipient were already linked in the previous round; whether proponent and recipient were indirectly linked through other agents in the previous round; whether the proponent attempted to establish a link with the recipient in the previous round but failed, finally whether the recipient attempted to establish a link with the recipient in the previous round but failed.[7]

We identify two main drives to link formation: best response behaviour and attempt to coordinate on an efficient architecture. Rationality requires agents to respond to the present network by establishing direct links to those who have a larger number of connections and not to propose a direct link to those that they can otherwise reach through indirect connections. Our estimates show that the likelihood of proposing a link is significantly affected by the number of links that the recipient has in the previous round. Moreover we find strong evidence of the fact that whenever the proponent and the recipient are indirectly linked in the previous round, a link proposal is less likely. Hence costly link formation is indeed directed to increase the profits that accrue to agents when they establish connections to those nodes that they are not able to reach otherwise.

The second drive to link formation is in the attempt to coordinate on an efficient architecture: participants try to avoid unmatched proposals by using the information that they have about subjects that have proposed to them in previous rounds and by relying on repeated interaction. The likelihood of *i* proposing a link to *j* increases when *i* and *j* were linked in the previous round, it increases when *j* has proposed a link to *i* and has failed the connection in the previous round. Subjects are also more likely to propose again to recipients that did not reciprocate in the past. As a result agents with many links are also more likely to propose links in the future. Not surprisingly we find that agents are less likely to propose links when the cost of link formation is higher.

Tables 5c and 5d present the results for our probit model respectively without and with demographic variables. The regression with demographics shows that the gender of experimental subjects does not affect the likelihood of link proposals. On the other hand, our regional dummies are significant with subjects from the south of Italy being less likely to form connections.

Table 5c Determinants of Individual Behaviour. Random effects probit model

Dependent Variable: *i* proposes a link to *j*	Coefficient	Std. Error	P-Value
links of *j* in *(t-1)*	0.07	0.01	**0.000**
links of *I* in *(t-1)*	0.10	0.02	**0.000**
i and *j* linked in *(t-1)*	0.21	0.07	**0.004**
j failed with *i* in *(t-1)*	0.39	0.05	**0.000**
i failed with *j* in *(t-1)*	0.26	0.05	**0.000**
i and *j* indirectly linked in *(t-1)*	–0.17	0.06	**0.006**
unitary cost of link	–0.05	0.00	**0.009**
constant	0.11	0.22	0.62
Number of observations: 6720			
Log-likelihood: –4181.4676			

Table 5d Determinants of Individual Behaviour. Random effects probit model with demographics

Dependent Variable: *i* proposes a link to *j*	Coefficient	Std. Error	P-Value
links of *j* in *(t-1)*	0.06	0.01	**0.000**
links of *I* in *(t-1)*	0.10	0.02	**0.000**
i and *j* linked in *(t-1)*	0.21	0.07	**0.003**
j failed with *i* in *(t-1)*	0.39	0.05	**0.000**
i failed with *j* in *(t-1)*	0.26	0.05	**0.000**
i and *j* indirectly linked in *(t-1)*	−0.17	0.06	**0.007**
unitary cost of link	−0.01	0.00	**0.015**
constant	–0.01	0.24	0.970
Female	0.03	0.07	0.709
North and Centre	0.15	0.07	**0.002**
Number of observations: 6720			
Log-likelihood: −4178.8907			

4 Conclusions

In this paper we explore network formation behaviour in a laboratory experiment. Interesting insights stem from both a micro and a macro level analysis.

We present two treatments. In the first one (low cost), any minimally connected network is both pairwise stable and efficient. In the second one (high cost), the theoretical prediction for a stable architecture is indeed unique (the empty graph) but inefficient: any minimally connected graph yields higher aggregate payoffs than the empty graph. For the low cost treatment, in accordance with the existing literature, we find that convergence to a stable network architecture is problematic because of the coordination problem caused by the multiplicity of equilibria. Quite interestingly, we also find lack of convergence in the high cost treatment, despite the fact that the equilibrium here is unique and does not require much coordination. We attribute this finding to the fact that subjects aimed at the efficient network, which is – again – not unique.

Despite lack of convergence, we detect a tendency to inclusion in the low cost treatment: the number of isolated nodes in the network of social ties decreases rapidly over time. In the low cost treatment this happens at the cost of greater redundancy: the number of redundant links increases over time. In the high cost treatment, we detect a tendency to minimality: the number of redundant links decreases over time. Symmetrically, this occurs at the cost of lower inclusion: the number of isolated nodes does not seem to decrease here as in the case of the low cost treatment.

The most commonly observed deviations from stable networks are: overconnectedness and the fact that minimally connected graph reached earlier on in the session are later departed from. A possible explanation for overconnectedness is that, due to multiplicity of equilibria, subjects try to cope with strategic uncertainty by forming redundant links as a form of insurance. Some aversion to inertia may explain the latter phenomenon.

From a micro perspective, we detect two main drives to link formation: best-response behaviour and attempt to coordinate on an efficient architecture. We find that, as predicted by a rational best-response behaviour, subjects are more likely to propose links to those who have a larger number of connections and are less likely to propose a link to those that they expect to be able to reach indirectly through the ties established by others. At the same time participants attempt to avoid miscoordination by proposing to those whose proposals they have received in the past and by relying on inertia and repeated interaction.

Appendix 1: Instructions (English Translation)

Welcome

This is an experiment on the formation of links among different subjects. If you make good choices you will be able to earn a sum of money that will be paid to you in cash immediately at the end of this session.

You are one of the six participants to this experiment; at the very beginning the computer will randomly assign you an initial budget (equal across participants). Also, the computer will randomly assign you an icon (**Dropper**, **Radio**, **Cube**, **Floppy**, **Hand lens**, **Hour glass**) that will to identify you throughout the experiment and will assign you an initial budget (equal across participants). The icon that identifies you is circled in red on your screen.

The experiment consists of a random number of rounds: there will be at least 15 rounds, after which a lottery administered by the computer will determine whether there is any further round or the experiment is over.

Each participant to this experiment represents a node. At the beginning of the experiment all nodes are isolated. In each round the computer will ask you whether you want to propose any link and to whom. You can propose 0, 1 or more links. *The computer will collect the proposals from all participants and will activate only the links which are desired by both subjects involved (reciprocated proposals)*.

Your screen will show the graph of active links. The box at the bottom right of your screen will show you who has proposed you a link in the previous round and whom you have not reciprocated.

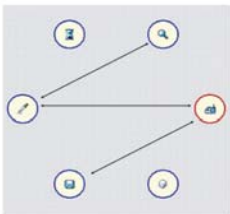
Each link that you manage to activate has a cost (equal across participants) that is indicated on the screen. At each round the computer may reject your link proposals if they entail an expenditure that is higher than your budget for that round.

Your revenues in each round are automatically computed and are given by the product by the revenue per node (equal across subjects and indicated on your screen) and the number of nodes that you manage to reach both through your direct links and the links activated by other participants.

Computing costs and revenues

Example: subject **Radio** is directly linked to **Floppy** and **Dropper** and indirectly, that is through **Dropper**, to **Hand lens**.

Unitary revenue: 10
Cost of each connection: 3



The profit of **Radio** is:

total revenues– total costs

total revenues = number of nodes reached (directly and indirectly) x revenue per node = 3 x 10 = 30

total costs = direct connections x cost of each connection = 2 x 3 = 6

profit = 24

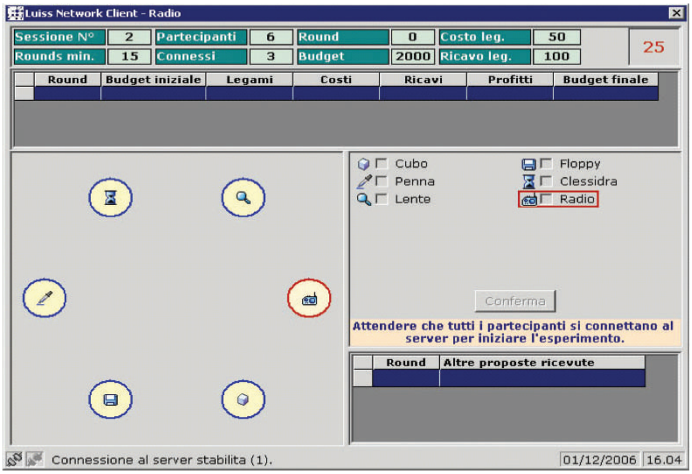
In each round the computer will work out your profit and will display it on your screen. The overall profit from the experiment is given by the sum of your profits in all rounds. At the end of the experiment you will be paid in cash an amount equivalent to the 10% of your total profit.

More in detail

At the beginning of the experiment please wait for instructions from the experimenters without touching any key.

When the experimenter will ask you to do so, please double-click only once on the "Network Client" icon on your desktop.

The following screen will appear:



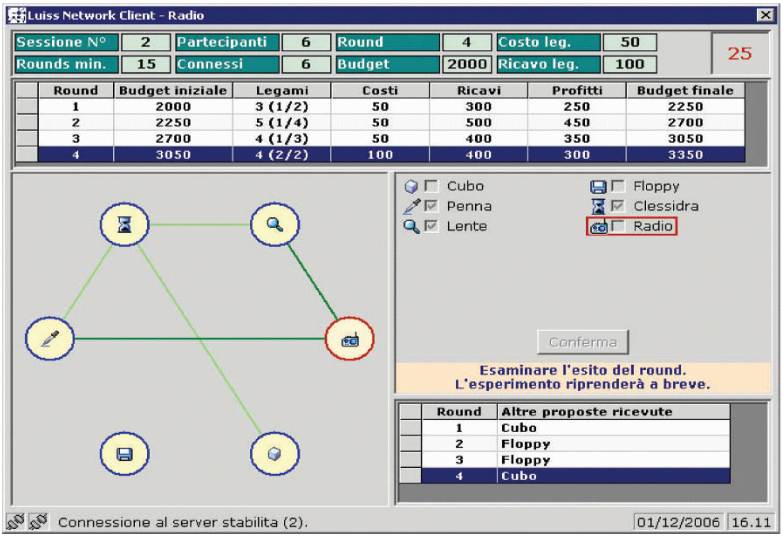
The screen gives you all the information regarding the round that you are about to play.

Be careful: each round has a maximum time duration given by the number of seconds indicated in red at the top-right of your screen. If you have not managed to make your choice by then, the computer will immediately proceed to the next round.

Your screen shows all data relative to the current round (available budget, costs and revenues) as well as the results that you have obtained from each of the previous rounds.

At the end of each round, the graph will show the links which have been activated by you and the other participants (as shown above). Moreover the table that summarises your performance in the current round will be updated. You will have the possibility to review the situation of previous rounds by clicking on the corresponding bar in the same table. The table at the bottom right of your screen gives you additional information on proposals that you have received but not matched in the previous rounds.

When the message "Round is now active" appears at the bottom of your screen, you can make your choice by ticking the boxes corresponding to the icons that you want to propose a link to. When you are done, press "Confirm". When all participants have confirmed their choices, the computer will show the results of the round on the screen.



You will be advised of the beginning of a new round by a "New Round" message. Be careful: after the 15th round, red and green lights will flash on the screen. If the lights stop as green, you will play another round; if they stop as red, the experiment is over.

It is very important that you make choices independently and that you do not communicate with other participants during the experimental session.

At the end of the last round the experiment is over and you will be paid in cash for a sum corresponding to your profit during the course of the whole experiment.

For any problem, please contact the experimenters.

Enjoy.

May 2006

Acknowledgements We would like to thank Anna Conte for helpful support in analysing the data. We are grateful to Alessandra Cassar, John Hey and other participants to FUR 2006 in Rome for useful discussion and suggestions. This paper benefited from comments by participants at ESA 2004 where a previous version was presented. Financial support by DIPTEA – LUISS, Rome and by EC FP6 Integrated Project 001907 (DELIS) is gratefully acknowledged.

Notes

[1]The effect of network membership and architecture on outcomes has also been studied through laboratory experiments. See for example, Cassar (2007) for an experimental study on network effects on cooperation and coordination, and Cassar, Crowley, and Wydick (2007) for a field experiment on social network effects on trust. For a thorough review on experiments with exogenous networks, see Kosfeld (2004).

[2]We believe that such network formation protocol is not only more realistic, but also that it may address some of the concerns on lack of convergence to an efficient architecture because of inequality aversion expressed by the literature (see, for example, Falk and Kosfeld (2003)).

[3]The software utilised for the experiment has been developed by InformaRoma. A special thanks goes to Andrea Lombardo.

[4]At the end of round 15 (and of each additional round after that), a lottery administered by the computer decided if an additional round had to be played. The probability of new rounds was fixed at 50%. The lottery was visualised on participants' screens as two flashing buttons, one red (with a NO sign) and one green (with a YES sign).

[5]In this setting we want to avoid any salient coordination device that induces coordination on a particular network. In the pilot for this experiment (see Di Cagno and Sciubba (2005)) we labeled participants with A, B, C, D, E, F and we found that the alphabetical ordering was explaining most of the networking decisions. See also Bernasconi and Galizzi (2005) and Falk and Kosfeld (2003).

[6]It is easy to check that average profits for the chain are also equal to 350.

[7]Recall that in our setting each subject is informed about those players who have made link proposals to them in previous rounds and have not been reciprocated.

References

Bala, V., & Goyal, S. (2000). A non-cooperative model of network formation. *Econometrica, 68*(5), 1181–1229

Bernasconi, M., & Galizzi, M. (2005). *Coordination in networks formation: Experimental evidence on learning and salience* (FEEM Working Paper No. 107)

Berninghaus, S. K., Ott, M., & Vogt, B. (2004). On networks and stars: Recent results in network experiments. (Mimeo)

Callander, S., & Plott, C. (2005). Principles of network development and evolution: An experimental study. *Journal of Public Economics, 89*(8), 1469–1495

Cassar, A. (2007). Coordination and cooperation in local, random and small world networks: Experimental evidence. *Games and Economic Behaviour, 58*(2), 209–230

Cassar, A., Crowley, L., & Wydick, B. (2007). The effects of social capital on group loan repayment: Evidence from field experiment. *Economic Journal, 117*(517), F85–F106

Deck, C., & Johnson, C. (2004). Link bidding in laboratory network. *Review of Economic Design, 8*(4), 359–372

Di Cagno, D., & Sciubba, E. (2005). *Are individuals profit maximizing in network formation? Some experimental evidence* (DPTEA Working Paper No. 138)

Falk, A., & Kosfeld, M. (2003). *It's all about connections: Evidence on network formation* (IEW Working Paper No. 21)

Goeree, J. K., Riedl, A., & Ule, A. (2005). *In search of stars: Network formation among heterogeneous agents* (IZA Working Paper No. 1754)

Goyal, S., & Joshi, S. (2006). Unequal connections. *International Journal of Game Theory, 34*(3), 319–349

Jackson, M. O., & Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory, 71*(1), 44–74

Kosfeld, M. (2004). Economic networks in the laboratory: A survey. *Review of Network Economics, 3*(1), 20–41

Vanin, P. (2002). Network formation in the lab: A pilot experiment. (Mimeo)