Discussion Papers No. 409, March 2005
Statistics Norway, Research Department

John K. Dagsvik

Choice under Uncertainty and Bounded Rationality

Abstract:
This paper develops a theory for probabilistic models for risky choices that can be viewed as an extension of the expected utility theory to account for bounded rationality. One probabilistic version of the *Archimedean Axiom* and two versions of the *Independence Axiom* are proposed. In addition, additional axioms are proposed of which one is Luce's *Independence from Irrelevant Alternatives*. It is demonstrated that different combinations of the axioms yield different characterizations of the probabilities for choosing the respective risky prospects.

Keywords: Random tastes, bounded rationality, independence from irrelevant alternatives, choice among lotteries, probabilistic choice for uncertain outcomes.

JEL classification: C25, D11, D81

Acknowledgement: I have benefited from comments from Jørgen Weibull, Steinar Strøm and participants in workshops in Statistics Norway. Thanks to Marina Rybalka and Weizhen Zhu who pointed out a serious error, and to Anne Skoglund for excellent word processing assistance.

Address: John K. Dagsvik, Statistics Norway, Research Department. E-mail: john.dagsvik@ssb.no

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
NO-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1. Introduction

In the standard theory of decision under uncertainty it is assumed that the agent's preference functional is deterministic. This assumption is also maintained in most of the recent theoretical and empirical literature. It has been recognized for some time, however, that even in seemingly identical repetitions (replications) of the same choice setting the decision maker often makes inconsistent choices. This means that the deterministic theory cannot be directly applied in an empirical context unless some additional stochastic "error" is introduced. As Fishburn (1976, 1978), Hey (1995), Carbone (1997), Loomes and Sugden (1995), (1998) and Starmer (2000) discuss, this raises the question of how the axiomatization of theories for choice under uncertainty should be extended to accommodate stochastic error.

This paper proposes an axiomatic foundation of probabilistic models for risky choice experiments that may be viewed as a generalization of the von Neumann-Morgenstern expected utility theory. This setting means that the agent's choice behavior in replications of choice settings (with uncertain outcomes) is assumed to be governed by a probability mechanism. By now there is a huge literature on stochastic choice models with certain outcomes. In fact, it was empirical observations of inconsistencies, dating back to Thurstone (1927a,b), that led to the study of probabilistic theories in the first place. Thurstone argued that one reason for observed inconsistent choice behavior is bounded rationality in the sense that the agent is viewed as having difficulties with assessing the precise value (to him) of the choice objects. While probabilistic models for certain outcomes have been studied and applied extensively in psychology and economics it seems that there has been little interest for developing corresponding models for choice with uncertain outcomes. (For a summary of models with uncertain outcomes, see Fishburn (1998) and Starmer (2000, Section 6.2).) This is somewhat curious since one would expect that if an agent has problems with rank ordering alternatives with certain outcomes he would certainly find it difficult to choose among lotteries. The importance of developing theoretically justified stochastic choice models for uncertain outcomes has been articulated by Harless and Camerer (1994), and Hey and Orme (1994). For example, Hey and Orme, p.p. 1321-1322, summarize their view as follows:

> "....., we are tempted to conclude by saying that our study indicates that behavior can be reasonably well modelled (to what might be termed a 'reasonable approximation') as 'Expected utility plus noise'. Perhaps we should now spend some time thinking about the noise, rather than about even more alternatives to expected utility?"

One of the axioms we propose in this paper can be viewed as a probabilistic version of the so called *Archimedean Axiom* and two of the axioms can be viewed as probabilistic versions of the *Independence Axiom* in the von Neumann-Morgenstern theory of expected utility. These probabilistic versions extend the basic von Neumann-Morgenstern axioms in the following sense: While the

Archimedean and *Independence* may not necessarily hold in a single choice experiment, the probabilistic versions state that they will hold in an aggregate sense (to be made precise below) when the agent participates in a large number of replications of a choice experiment. The intuition is that the agent may be boundedly rational and make errors when he evaluates the value to him of the respective choice alternatives (strategies) in each single replication of the experiment but, on average (across replications of the experiment), show no systematic departure from the von Neumann-Morgenstern type of axioms. Alternatively, the probabilistic axioms may also be conveniently interpreted in the context of an observationally homogeneous *population* of agents that face the same choice experiment: While each agent's behavior is allowed to deviate from the von Neumann-Morgenstern axioms, the "aggregate" behavior in the population is assumed to be consistent with these axioms. The latter type of interpretation is the most common one within the theory of discrete choice (see for example McFadden, 1981, 1984).

We first consider the case with comparisons between two lotteries. We demonstrate that different combinations of the probabilistic Archimedean and Independence axioms combined with other additional axioms, imply particular characterizations of the probabilities for choice among risky prospects as a function of the lottery outcome probabilities (given the choice).

Subsequently, we consider the multinomial case where the agent faces a choice set of several lotteries. In this case we apply different combinations of the axioms mentioned above together with Luce's Choice Axiom: "Independence from Irrelevant Alternatives" (IIA). In this context, IIA yields a Luce model where the utility of a lottery is a general function of the lottery outcome probabilities associated with this lottery.

The choice probabilities implied by the proposed axioms are essential for establishing the link between theory and the corresponding empirical model. More precisely, the agents' choices among lotteries are, from a statistical point of view, outcomes of a multinomial experiment with probabilities that are the respective choice probabilities mentioned above. Accordingly, when the structure of the choice probabilities has been obtained one can, in the context of empirical analysis, apply standard inference methods such as maximum likelihood estimation procedures and likelihood ratio tests, for example.

The first work on stochastic models for choice among lotteries occured in the sixties. Becker et al. (1963b) proposed a probabilistic model for choice among lotteries which they called a *Luce Model for Wagers*. Luce and Suppes (1965) considered a special case of the Luce model for wagers which they called the *Strict Expected Utility Model*. However, neither these authors nor more recent contributions discuss the issue of deriving a stochastic model from axiomatization. To the best of our knowledge the only contribution that provide a model that is founded on an axiomatic basis is Fishburn (1978) who develops the *incremental expected utility advantage* model. His model does, however, not contain the expected utility model as a special case, although it can be approximated by an incremental expected utility advantage model. As pointed out by Fishburn (1978), pp. 635-636, the incremental advantage model seems extreme since it implies that there is a positive probability of choosing a \$ 1 in a choice between \$ 1 for certain or a gamble that yields \$ 10 000 with probability .999 or \$ 0 with probability .001.

The paper is organized as follows. In the next section we present some basic concepts and notation. In Section 3 we discuss different types of axiomatizations and characterizations for binary choice models. In Section 4 we consider axiomatizations and their implications for the multinomial choice setting. In Section 5 we specialize to the case with monetary rewards, and in Section 6 we discuss a random utility representation. Finally, we discuss an example in Section 7.

2. Preliminaries

The aim of this section is to introduce axioms that enable us to characterize choice among lotteries when there is some randomness in the agent's choice behavior in the sense that if he faces several replications of a specific choice experiment he may choose different lotteries each time. The reason for this type of inconsistent behavior may be, as mentioned above, that he may have difficulty with evaluating the proper value (to him) of the respective lotteries.

Let X denote the set that indexes the set of outcomes, which is assumed to be finite and contains m outcomes, i.e. $X \equiv \{1, 2, ..., m\}$. In the following we shall assume, as is customary, that the agent's information about the chances of the different realizations of lottery s can be represented by lottery outcome probabilities;

$$\mathbf{g}_s := \left(g_s(1), g_s(2), ..., g_s(m)\right)$$

where $g_s(k)$ is the probability of outcome k, $k \in X$, if lottery s is chosen. Let S denote the set of simple probability measures on the algebra of all subsets of the set of outcomes. Recall that by a *preference relation* it is meant a binary relation, $\succeq$, on S that is (i) *complete*, i.e. for all $\mathbf{g}_r$, $\mathbf{g}_s \in S$ either $\mathbf{g}_r \succeq \mathbf{g}_s$ or $\mathbf{g}_s \succeq \mathbf{g}_r$, and (ii) *transitive*, i.e. for all $\mathbf{g}_r$, $\mathbf{g}_s$, $\mathbf{g}_t$, in S, $\mathbf{g}_r \succeq \mathbf{g}_s$ and $\mathbf{g}_s \succeq \mathbf{g}_t$ implies $\mathbf{g}_r \succeq \mathbf{g}_t$. A real-valued function $L(\mathbf{g}_s)$ on S *represents* $\succeq$ if for all $\mathbf{g}_r$, $\mathbf{g}_s \in S$, $\mathbf{g}_r \succeq \mathbf{g}_s$, if and only if $L(\mathbf{g}_r) \geq L(\mathbf{g}_s)$. Let $\mathfrak{B}$ be the family of *finite* subsets of S that contain at least two elements.

Consider now the following choice setting: The agent faces n replications of a choice experiment in which a set B of lotteries, $B \in \mathfrak{B}$, is presented in each replication. We assume that there is no learning. Since there is an element of randomness in the agent's choice behavior he may choose different lotteries in different replications. We assume that the agent's choices in different replications are stochastically independent. Let $P_B(\mathbf{g}_s)$, $\mathbf{g}_s \in B$, be the probability that $\mathbf{g}_s$ is the most preferred vector of lottery outcome probabilities in B. Let $P(\mathbf{g}_r, \mathbf{g}_s)$ be the probability that lottery $\mathbf{g}_r$ is chosen over $\mathbf{g}_s$, i.e., $P(\mathbf{g}_r,\mathbf{g}_s) \equiv P_{\{\mathbf{g}_r,\mathbf{g}_s\}}(\mathbf{g}_r)$. It then follows that $P(\mathbf{g}_r,\mathbf{g}_s) > P(\mathbf{g}_s,\mathbf{g}_r)$ if and only if $P(\mathbf{g}_r,\mathbf{g}_s) > 0.5$. The argument above provides a motivation for the following definition:

Definition 1

For $\boldsymbol{g}_r, \boldsymbol{g}_s \in S$, lottery $\boldsymbol{g}_r$ is said to be strictly preferred to $\boldsymbol{g}_s$ in the aggregate sense, if and only if $P(\boldsymbol{g}_r,\boldsymbol{g}_s) > 0.5$. If $P(\boldsymbol{g}_r,\boldsymbol{g}_s) = 0.5$, then $\boldsymbol{g}_r$ is, in the aggregate sense, indifferent to $\boldsymbol{g}_s$.

Thus, Definition 1 introduces a binary relation, $\succeq$, where $\mathbf{g}_r \succ \mathbf{g}_s$ means that $\mathbf{g}_r$ is strictly preferred to $\mathbf{g}_s$ (in the aggregate sense), while $\mathbf{g}_r \sim \mathbf{g}_s$ means that $\mathbf{g}_r$ is indifferent to $\mathbf{g}_s$. Note, however, that the relation is not necessarily a *preference relation*. The reason for this is that the binary relation $\succeq$ is *not* necessarily transitive. That is, for $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 \in S$ the statement; $P(\mathbf{g}_1,\mathbf{g}_2) \geq 0.5$ and $P(\mathbf{g}_2,\mathbf{g}_3) \geq 0.5$ imply $P(\mathbf{g}_1,\mathbf{g}_3) \geq 0.5$, is not necessarily true.

Let $\mathbf{g}_1, \mathbf{g}_2 \in S$. The mixed lottery, $\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_2$, $\alpha \in [0,1]$, is a lottery in S yielding the probability $\alpha \mathbf{g}_1(k) + (1-\alpha)\mathbf{g}_2(k)$, of outcome k, $k \in X$. Here we assume that the agents perceive the lotteries $\alpha\beta \mathbf{g}_1 + (1-\alpha\beta)\mathbf{g}_2$ and $\beta\left[\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_2\right] + (1-\beta)\mathbf{g}_2$, $\alpha, \beta \in [0,1]$, as equivalent. This property is known as the axiom of reduction of compound lotteries, cf. Luce and Raiffa (1957).

For sets, $A, B \in \mathfrak{B}$ such that $A \subseteq B$, let

$$P_B(A) \equiv \sum_{\mathbf{g}_s \in A} P_B(\mathbf{g}_s).$$

The interpretation is that $P_B(A)$ is the probability that the agent shall choose a lottery within A when B is the choice set.

3. Binary choice probabilities between lotteries

In this section we shall extend the von Neumann-Morgenstern expected utility theory to a corresponding probabilistic theory, in the sense discussed above. We shall in this section only discuss the case of binary choice. We shall next introduce a set of behavioral axioms which will lead to different types of characterizations of the binary choice probabilities. The purpose of the first axiom is to impose necessary and sufficient conditions to insure that the binary relation given in Definition 1 is a preference relation.

Axiom 1 (*Weak Stochastic Transitivity*)

Let $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3 \in S$ *. The binary choice probabilities satisfy*

(i) *Weak stochastic transitivity; if* $P(\boldsymbol{g}_1, \boldsymbol{g}_2) \geq \frac{1}{2}$ *and* $P(\boldsymbol{g}_2, \boldsymbol{g}_3) \geq \frac{1}{2}$*, then* $P(\boldsymbol{g}_1, \boldsymbol{g}_3) \geq \frac{1}{2}$.

(ii) *The Balance condition;* $P(\boldsymbol{g}_1, \boldsymbol{g}_2) + P(\boldsymbol{g}_2, \boldsymbol{g}_1) = 1$.

Note that the *Balance condition* is equivalent to *completeness*. It follows immediately that the binary relation given in Definition 1 is a *preference relation* provided it satisfies Axiom 1.

Axiom 2 (*Archimedean*)

For all $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3 \in S$ *, if*

$$P(\boldsymbol{g}_1, \boldsymbol{g}_2) > \frac{1}{2} \quad \text{and} \quad P(\boldsymbol{g}_2, \boldsymbol{g}_3) > \frac{1}{2}$$

then there exist $\alpha, \beta \in (0,1)$ *such that*

$$P(\alpha \boldsymbol{g}_1 + (1-\alpha)\boldsymbol{g}_3, \boldsymbol{g}_2) > \frac{1}{2} \quad \text{and} \quad P(\boldsymbol{g}_2, \beta \boldsymbol{g}_1 + (1-\beta)\boldsymbol{g}_3) > \frac{1}{2}.$$

Axiom 2 is a probabilistic version of the *Archimedean Axiom* in the von Neumann-Morgenstern expected utility theory since, by Definition 1, it is equivalent to the following statement: If $\mathbf{g}_1 \succ \mathbf{g}_2$ and $\mathbf{g}_2 \succ \mathbf{g}_3$, then there exists $\alpha, \beta \in (0,1)$ such that

$$\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_3 \succ \mathbf{g}_2$$

and

$$\mathbf{g}_2 \succ \beta \mathbf{g}_1 + (1-\beta)\mathbf{g}_3,$$

cf. Karni and Schmeidler (1991), p. 1769. Note that Axiom 2 is weaker than the assumption that $P(\mathbf{g}_r, \mathbf{g}_s)$ is continuous. This is so because if $P(\mathbf{g}_r, \mathbf{g}_s)$ is continuous in $(\mathbf{g}_r, \mathbf{g}_s)$, then whenever $P(\mathbf{g}_1, \mathbf{g}_2) > 1/2$ and $P(\mathbf{g}_2, \mathbf{g}_3) > 1/2$, continuity implies that

$$P(\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_3, \mathbf{g}_2) > 1/2$$

and

$$P(\mathbf{g}_2, \beta \mathbf{g}_1 + (1-\beta)\mathbf{g}_3) > 1/2$$

for suitable $\alpha,\ \beta \in (0,1)$.

Axiom 3 (*Independence*)

For all $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3 \in S$ *, and all* $\alpha \in [0,1]$ *, if*

$$P(\boldsymbol{g}_1, \boldsymbol{g}_2) \geq \frac{1}{2}$$

then

$$P(\alpha \boldsymbol{g}_1 + (1-\alpha)\boldsymbol{g}_3, \alpha \boldsymbol{g}_2 + (1-\alpha)\boldsymbol{g}_3) \geq \frac{1}{2}.$$

Axiom 3 is a probabilistic version of the *Independence Axiom* in the von Neumann-Morgenstern expected utility theory, because it is equivalent to the statement: If $\mathbf{g}_1 \succ \mathbf{g}_2$, then $\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_3 \succ \alpha \mathbf{g}_2 + (1-\alpha)\mathbf{g}_3$, cf. Karni and Schmeidler (1991), p. 1769.

Theorem 1 (*von Neumann-Morgenstern*)

Let $\succeq$ *be a binary relation. The following two conditions are equivalent:*

(i) $\succeq$ *is a preference relation satisfying Axioms 2 and 3.*

(ii) There exists a function, $u : X \to R$ *, that is unique up to a positive affine transformation such that the function* $V : S \to R$ *defined by*

$$V(\boldsymbol{g}) = \sum_{k \in X} u(k) g(k)$$

represents the preference relation.

Theorem 1 is the von Neumann-Morgenstern Expected Utility Theorem, cf. Karni and Schmeidler (1991), pp. 1769-70.

Recall that we cannot apply the result of Theorem 1 directly in our context since the binary relation of Definition 1 is not necessarily a preference relation.

Since the binary relation given in Definition 1 is a preference relation when it satisfies Axiom 1, the next corollary follows.

Corollary 1

Assume that Axioms 1, 2 and 3 hold. Then for $\boldsymbol{g}_1, \boldsymbol{g}_2 \in S$ *,*

$$P(\boldsymbol{g}_1,\boldsymbol{g}_2) \geq \frac{1}{2} \Leftrightarrow V(\boldsymbol{g}_1) \geq V(\boldsymbol{g}_2).$$

Moreover, if either antecedent inequality is strict so is the conclusion.

Even if the binary relation given in Definition 1 satisfies Axioms 1, 2 and 3, we would still not be able to specify choice probabilities. We would at most be able to ascertain whether or not $\mathbf{g}_r$ is preferred to $\mathbf{g}_s$ (say) in the aggregate sense. Consequently, we need to provide additional theoretical building blocks so as to be able to ascertain precisely how the choice probabilities $\{P(\mathbf{g}_r, \mathbf{g}_s)\}$ can be represented by the lottery outcome probabilities $\mathbf{g}_r$ and $\mathbf{g}_s$. This is crucial for establishing a link between the theoretical concepts introduced above and a model that is applicable for empirical modeling and analysis. The next axiom is useful in this respect.

Axiom 4 (*Order-independence*)

For all $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3 \in S$

$$P(\boldsymbol{g}_1,\boldsymbol{g}_2) \geq \frac{1}{2} \text{ if and only if } P(\boldsymbol{g}_1,\boldsymbol{g}_3) \geq P(\boldsymbol{g}_2,\boldsymbol{g}_3).$$

Axiom 4 is a special case of what is called the *order-independence* condition, see Suppes et al. (1989), p.p. 411-412. The intuition is that if $\mathbf{g}_1$ is chosen more frequently than $\mathbf{g}_2$, then the fraction of times $\mathbf{g}_1$ is preferred over $\mathbf{g}_3$ is higher than the fraction of times $\mathbf{g}_2$ is preferred over $\mathbf{g}_3$.

Theorem 2

For all $\boldsymbol{g}_1, \boldsymbol{g}_2 \in S$*, Axioms 2 to 4 hold if and only if*

(3.1) $$P(\boldsymbol{g}_1,\boldsymbol{g}_2) = F\left(V(\boldsymbol{g}_1), V(\boldsymbol{g}_2)\right)$$

where

$$V(\boldsymbol{g}_s) = \sum_{k \in X} u(k) g_s(k),$$

$F: R^2 \rightarrow (0,1)$ *is a function that is strictly increasing in its first argument and strictly decreasing in the second, and* $u: X \rightarrow R$ *is a function that is unique up to a positive linear transformation.*

The proof of Theorem 2 is given in the Appendix.

When $\mathrm{P}(\mathbf{g}_1,\mathbf{g}_2)$ can be represented as a function $\mathrm{F}(\mathrm{f}(\mathbf{g}_1),\mathrm{f}(\mathbf{g}_2))$, for some suitable scalar function f defined on S, and F is strictly increasing in its first argument and strictly decreasing in the second, the choice probabilities is said to be *simply scalable,* cf. Suppes et al. (1989), p. 410. They prove that Axiom 4 is equivalent to *simple scalability.* The representation (3.1) seems to be the weakest possible representation for choice under uncertainty one can think of. It would include any kind of probabilistic binary non-expected utility model since the function f is allowed to be completely general. Despite its generality simple scalability is violated in some choice contexts, see for example Problem 2 in Suppes et al. (1989), p. 413.

Axiom 5

Let $\boldsymbol{g}_1,\boldsymbol{g}_2,\boldsymbol{g}_3,\boldsymbol{g}_4 \in S$ *. The binary choice probabilities satisfy*

(i) *The Quadruple condition;* $P(\boldsymbol{g}_1,\boldsymbol{g}_2) \geq P(\boldsymbol{g}_3,\boldsymbol{g}_4)$ *if and only if* $P(\boldsymbol{g}_1,\boldsymbol{g}_3) \geq P(\boldsymbol{g}_2,\boldsymbol{g}_4)$*. Moreover, if either antecedent inequality is strict so is the conclusion.*

(ii) *Solvability; For any* $y \in (0,1)$ *and any* $\boldsymbol{g}_1,\boldsymbol{g}_2,\boldsymbol{g}_3 \in S$ *satisfying* $P(\boldsymbol{g}_1,\boldsymbol{g}_2) \geq y \geq P(\boldsymbol{g}_1,\boldsymbol{g}_3)$ *, there exists a* $\boldsymbol{g} \in S$ *such that* $P(\boldsymbol{g}_1,\boldsymbol{g}) = y$ *.*

(iii) *The Balance condition;* $P(\boldsymbol{g}_1,\boldsymbol{g}_2) + P(\boldsymbol{g}_2,\boldsymbol{g}_1) = 1$ *.*

Axiom 5 is due to Debreu (1958). The intuition of the Quadruple condition is related to the following example where the binary choice probabilities have the form

$$\mathrm{P}(\mathbf{g}_1,\mathbf{g}_2) = \mathrm{G}(\mathrm{f}(\mathbf{g}_1) - \mathrm{f}(\mathbf{g}_2))$$

where G is a strictly increasing cumulative distribution function on R, and f is a suitable mapping from S to R. Clearly, the choice model in this example satisfies the Quadruple condition. The example shows that when the average value of some lottery outcome probabilities **g** is represented by a scale function, f(**g**), in such a way that the propensity to prefer lottery outcome probabilities $\mathbf{g}_1$ over $\mathbf{g}_2$ is a function of the “distance”, $\mathrm{f}(\mathbf{g}_1) - \mathrm{f}(\mathbf{g}_2)$, then the Quadruple condition must hold. Also the Solvability- and the Balance conditions are fairly intuitive. If G is continuous the Balance condition will also be fulfilled in the example above.

Theorem 3

Axiom 5 holds if and only if there exists a continuous and strictly increasing cumulative distribution function G with $G(x) + G(-x) = 1$*, and a mapping f from S to some interval I such that the binary choice probabilities can be represented as*

(3.2) $$P(\boldsymbol{g}_1,\boldsymbol{g}_2)=G\{f(\boldsymbol{g}_1)-f(\boldsymbol{g}_2)\}$$

for $\boldsymbol{g}_1,\boldsymbol{g}_2\in S$ *, where*

$$I=\{x: x=f(\boldsymbol{g}),\boldsymbol{g}\in S\}.$$

The mapping f is unique up to a linear transformation. The c.d.f. G is unique in the sense that if G_1 *and* G_0 *are two representations, then* $G_0(x)=G_1(ax)$*, where* $a>0$ *is a constant.*

The proof of Theorem 3 is given in the Appendix. Similar results are obtained in Falmagne (1985) and Suppes et al. (1989). For the sake of completeness we shall, however, give the proof in the Appendix.

Theorem 4

For all $\boldsymbol{g}_1,\boldsymbol{g}_2\in S$ *, Axioms 2, 3 and 5 hold if and only if*

(3.3) $$P(\boldsymbol{g}_1,\boldsymbol{g}_2)=G\{h(V(\boldsymbol{g}_1))-h(V(\boldsymbol{g}_2))\}$$

where

(3.4) $$V(\boldsymbol{g}_r)=\sum_{k\in X}u(k)\boldsymbol{g}_r(k),$$

G is a continuous and strictly increasing cumulative distribution function defined on R with $G(x)+G(-x)=1$*,* $h: R\to R$ *is strictly increasing and* $u: X\to R$*.* $G(\cdot)$ *and* $h(V(\cdot))$ *are unique in the sense that if* $G_0(\cdot)$ *and* $G_1(\cdot), h_0(V_0(\cdot))$ *and* $h_1(V_1(\cdot))$ *are two representations, then* $G_0(x)=G_1(ax)$ *where* $a>0$ *is a constant,* $V_1(\boldsymbol{g}_r)=b_1V_0(\boldsymbol{g}_r)+c_1$ *and* $h_1(b_1x+c_1)=b_2h_0(x)+c_2$ *where* $b_1>0$*,* $b_2>0$*,* c_1 *and* c_2 *are constants.*

The proof of Theorem 4 is given in the Appendix.

Remark

Note that the formulation in (3.2) is equivalent to

$$\mathrm{P}(\mathbf{g}_1,\mathbf{g}_2)=\tilde{\mathrm{G}}\left(\tilde{\mathrm{h}}(\mathrm{V}(\mathbf{g}_1))/\tilde{\mathrm{h}}(\mathrm{V}(\mathbf{g}_2))\right)$$

where $\tilde{\mathrm{G}}$ is a continuous and strictly increasing c.d.f. on R_+ and $\tilde{\mathrm{h}}$ is positive and strictly increasing. This follows immediately from (3.3) by defining $\tilde{\mathrm{G}}(\mathrm{x})=\mathrm{G}(\mathrm{e}^{\mathrm{x}})$ and $\log\tilde{\mathrm{h}}(\mathrm{x})=\mathrm{h}(\mathrm{x})$.

Axiom 6 (Strong independence)

For all $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_1^*, \boldsymbol{g}_2^*, \boldsymbol{g}_3 \in S$ *and all* $\alpha \in [0,1]$, *if*

$$P(\boldsymbol{g}_1, \boldsymbol{g}_2) \geq P(\boldsymbol{g}_1^*, \boldsymbol{g}_2^*),$$

then

$$P(\alpha \boldsymbol{g}_1 + (1-\alpha)\boldsymbol{g}_3, \alpha \boldsymbol{g}_2 + (1-\alpha)\boldsymbol{g}_3) \geq P(\alpha \boldsymbol{g}_1^* + (1-\alpha)\boldsymbol{g}_3, \alpha \boldsymbol{g}_2^* + (1-\alpha)\boldsymbol{g}_3).$$

In other words, Axiom 6 states that if the fraction of replications where $\mathbf{g}_1^*$ is chosen over $\mathbf{g}_2^*$ is less than or equal to the fraction of replications where $\mathbf{g}_1$ is chosen over $\mathbf{g}_2$, this inequality still holds when $\mathbf{g}_j$ is replaced by $\alpha \mathbf{g}_j + (1-\alpha)\mathbf{g}_3$ and $\mathbf{g}_j^*$ is replaced by $\alpha \mathbf{g}_j^* + (1-\alpha)\mathbf{g}_3$, for $j = 1, 2.$ Note that in Axiom 6 it is *not* claimed that $P(\mathbf{g}_1, \mathbf{g}_2)$ is equal to $P(\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_3, \alpha \mathbf{g}_2 + (1-\alpha)\mathbf{g}_3)$.

It follows that Axiom 6 implies Axiom 3. To realize this note that when $g_1^* = g_2^*$, then $P(\mathbf{g}_1^*, \mathbf{g}_2^*) = 1/2$ and

$$P(\alpha \mathbf{g}_1^* + (1-\alpha)\mathbf{g}_3, \alpha \mathbf{g}_2^* + (1-\alpha)\mathbf{g}_3) = 1/2.$$

Thus, it follows from this and Axiom 6 that when

$$P(\mathbf{g}_1, \mathbf{g}_2) \geq 1/2$$

then

$$P(\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_3, \alpha \mathbf{g}_2 + (1-\alpha)\mathbf{g}_3) \geq 1/2$$

which we recognize as the Independence Axiom.

Axiom 6 is stronger than Axiom 3, because it makes a statement that involves comparisons between the degree to which a lottery is chosen over a second to the degree to which a third is chosen over a fourth. It is this strengthening that enables us to derive strong functional form restrictions on the choice probabilities, to be considered next.

Theorem 5

Axioms 2, 5 and 6 hold if and only if the choice probabilities have the form as given in Theorem 4 with

(3.5) $$h(x) = \beta x + \kappa$$

where $\beta > 0$ *and* κ *are constants.*

The proof of Theorem 5 is given in the Appendix.

4. Multinomial choice probabilities between lotteries

A crucial building block for the extension of our theory to the multinomial case is the following axiom.

Axiom 7 (IIA)

For given $\boldsymbol{g}_s \in S$, $P(\boldsymbol{g}_s, \boldsymbol{g}_r) \in (0,1)$ *for all* $\boldsymbol{g}_r \in S$. *Furthermore, for* $g_s \in A \subseteq B, A, B \in \mathfrak{B}$, *then*

$$P_B(\boldsymbol{g}_s) = P_A(\boldsymbol{g}_s) P_B(A).$$

Axiom 7 was first proposed by Luce (1959) in the context of probabilistic choice with certain outcomes, and it is called *"Independence from Irrelevant Alternatives"* (IIA). As is well known, it represents a probabilistic version of rationality in the following sense: Suppose the agent faces a set B of feasible lotteries. One may view the agent's choice as if it takes place in two stages. In stage one he selects a subset from B, which contains the most attractive alternatives. In the second stage he chooses the most preferred alternative from this subset. In the second stage the alternatives outside the subset selected in stage one are *irrelevant*. Thus, rationality is associated with the property that the agent only takes into consideration the lotteries within the *presented* choice set. The probability that a particular set A (say) shall be chosen in the first stage is $P_B(A)$, and the probability that $\mathbf{g}_s$ is chosen (when alternatives in B\A are irrelevant) is $P_A(\mathbf{g}_s)$. Thus, $P_B(A)P_A(\mathbf{g}_s)$ is the final probability of choosing $\mathbf{g}_s$. As indicated above, the crucial point here is that $P_A(\mathbf{g}_s)$ is *independent* of alternatives outside A. For the sake of interpretation let J(B) denote the agent's chosen lottery from B. With this notation we can express IIA as

$$P_B(\mathbf{g}_s) = P(J(B) = \mathbf{g}_s) = P(J(B) \in A) P(J(A) = \mathbf{g}_s).$$

The conditional probability of choosing $\mathbf{g}_s$ given that the choice belongs to A, equals

$$P(J(B) = \mathbf{g}_s | J(B) \in A) = \frac{P(J(B) = \mathbf{g}_s)}{P(J(B) \in A)},$$

so that IIA can be rewritten as

$$\mathrm{P}\left(\mathrm{J(B)} = \mathbf{g}_s \,\middle|\, \mathrm{J(B)} \in \mathrm{A}\right) = \mathrm{P}\left(\mathrm{J(A)} = \mathbf{g}_s\right).$$

While $\mathrm{P}\left(\mathrm{J(A)} = \mathbf{g}_s\right)$ is the probability of choosing $\mathbf{g}_s$ from a given choice set A, the conditional probability

$$\mathrm{P}\left(\mathrm{J(B)} = \mathbf{g}_s \,\middle|\, \mathrm{J(B)} \in \mathrm{A}\right)$$

expresses the conditional probability of choosing $\mathbf{g}_s$ from a given choice set B, and given that the choice from B belongs to A. Clearly,

$$\mathrm{P}\left(\mathrm{J(B)} = \mathbf{g}_s \,\middle|\, \mathrm{J(B)} \in \mathrm{A}\right)$$

will in general be different from

$$\mathrm{P}\left(\mathrm{J(A)} = \mathbf{g}_s\right).$$

They only coincides when IIA holds.

Since Axiom 7 is a probabilistic statement it means that it represents probabilistic rationality in the sense that lotteries outside the second stage choice set A may matter in single choice experiments but will not affect average behavior.

It follows from Luce (1959) that Axiom 7 holds if and only if there exists representative scale values, $\mathrm{f}\left(\mathbf{g}_s\right)$, for some function f, such that

(4.1)
$$\mathrm{P_B}\left(\mathbf{g}_s\right) = \frac{\exp\left(\mathrm{f}\left(\mathbf{g}_s\right)\right)}{\sum_{\mathbf{g}_r \in \mathrm{B}} \exp\left(\mathrm{f}\left(\mathbf{g}_r\right)\right)},$$

provided $\mathrm{P}\left(\mathbf{g}_r, \mathbf{g}_s\right) \in (0,1)$ for all $\mathbf{g}_r, \mathbf{g}_s \in \mathrm{S}, \mathrm{B} \in \mathfrak{B}$. Thus, Axiom 7 implies that the relation given in Definition 1 is a preference relation. Hence, Axiom 4 or 5 are no longer needed, since they are implied by Axiom 7. Under IIA, the representation (4.1) is the weakest possible representation one can think of. It would include any kind of probabilistic non-expected utility model since the function f is allowed to be completely general.

Theorem 6

Assume that $P\left(\boldsymbol{g}_r, \boldsymbol{g}_s\right) \in (0,1)$ *for all* $\boldsymbol{g}_r, \boldsymbol{g}_s \in S$ *. Then for* $B \in \mathfrak{B}$ *, Axioms 2, 3 and 7 hold if and only if*

(4.2) $$P_B(\boldsymbol{g}_s)=\frac{exp\left(h\left(V(\boldsymbol{g}_s)\right)\right)}{\sum_{\boldsymbol{g}_r\in B} exp\left(h\left(V(\boldsymbol{g}_r)\right)\right)}$$

where

(4.3) $$V(\boldsymbol{g}_s)=\sum_{k\in X} u(k)g_s(k)$$

and $h:R\to R$ *is strictly increasing and* $u:X\to R$. *The function* $h(V(\cdot))$ *is unique in the sense that if* $h_0(V_0(\cdot))$ *and* $h_1(V_1(\cdot))$ *are two representations, then* $V_1(\boldsymbol{g}_r)=bV_0(\boldsymbol{g}_r)+c$ *and* $h_1(bx+c)=h_0(x)+d$ *where* $b>0$, *c and d are constants.*

The proof of Theorem 6 is given in the Appendix.

The choice model obtained in Theorem 6 is a special case of the *Luce model for wagers*, proposed by Becker et al. (1963b). They postulated that

(4.4) $$\mathrm{P_B}(\mathbf{g}_s)=\frac{\psi\left(\mathrm{V}(\mathbf{g}_s)\right)}{\sum_{\mathbf{g}_r\in \mathrm{B}}\psi\left(\mathrm{V}(\mathbf{g}_r)\right)}$$

where $\psi:\mathrm{R}\to\mathrm{R}_+$ is a strictly increasing mapping that is unique up to a multiplicative constant. By letting $\log\psi(x)=h(x)$ we realize that (4.4) is equivalent to (4.2).

Corollary 2

Assume that $P(\boldsymbol{g}_r,\boldsymbol{g}_s)\in(0,1)$ *for all* $\boldsymbol{g}_r,\boldsymbol{g}_s\in S$. *Then for* $B\in\mathfrak{B}$, *Axioms 2, 6 and 7 hold if and only if*

$$P_B(\boldsymbol{g}_s)=\frac{exp\left(V(\boldsymbol{g}_s)\right)}{\sum_{\boldsymbol{g}_r\in B} exp\left(V(\boldsymbol{g}_r)\right)}.$$

Proof:

Since Axiom 6 implies Axiom 3 it follows from Theorem 6 that (4.2) must hold. Consider the special case with $\mathrm{B}=\{\mathbf{g}_1,\mathbf{g}_2\}$. In this case Theorem 5 applies and implies (3.5). Without loss of generality we can set $\kappa=0$ and $\beta=1$ since κ cancels and β is absorbed in the utilities $\{u(k)\}$ in the expression for the choice probability.

Q.E.D.

5. Monetary rewards

The next two axioms we shall discuss are somewhat different from the previous ones in that we focus on applications where money is involved. Specifically, we now assume that the index set of outcomes is replaced by $X \times W$, where W is a finite set of *money amounts* that takes values in K. Thus, the lottery outcomes of the choice experiment consists of pairs $(j, w) \in X \times W$. The corresponding probability of outcome (j,w) given strategy s is denoted by $g_s(j, w)$. The utilities are now given as $\{u(j, w)\}$. Let S and $\mathfrak{B}$ be defined as in Section 2, suitably extended to the present setting. What distinguishes the present setting from the previous one is that one component (money) of the outcome is an *ordered* variable.

Analogous to (3.4) let

(5.1) $$V_\lambda(\mathbf{g}_s) = \sum_{(k,w)\in X\times W} u(k, \lambda w) g_s(k, w)$$

where λ is a positive real number.

Axiom 8

Let $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_1^*, \boldsymbol{g}_2^* \in S$, *and suppose that (3.3) hold (suitably extended to the present context), with h(·) and u(k,·) continuous and strictly increasing on R_+ for all k. Moreover, G is symmetric, continuous and strictly increasing. Then*

$$G\Big(h\big(V_1(\boldsymbol{g}_1)\big) - h\big(V_1(\boldsymbol{g}_2)\big)\Big) \le G\Big(h\big(V_1(\boldsymbol{g}_1^*)\big) - h\big(V_1(\boldsymbol{g}_1^*)\big)\Big)$$

if and only if

$$G\Big(h\big(V_\lambda(\boldsymbol{g}_1)\big) - h\big(V_\lambda(\boldsymbol{g}_2)\big)\Big) \le G\Big(h\big(V_\lambda(\boldsymbol{g}_1^*)\big) - h\big(V_\lambda(\boldsymbol{g}_1^*)\big)\Big)$$

for any $\lambda > 0$.

Axiom 8 means that if the fraction of individuals that prefer $\mathbf{g}_1$ over $\mathbf{g}_2$ is less than the fraction that prefers $\mathbf{g}_1^*$ over $\mathbf{g}_2^*$ then this inequality does not change if all the incomes are re-scaled by the same factor while the lottery outcome probabilities remain unchanged.

Before we state the next result we adopt the definition

$$\frac{x^\theta - 1}{\theta} = \log x$$

when $\theta = 0$. The function $(x^\theta - 1)/\theta$ will then be differentiable and strictly increasing for all θ.

Theorem 7

Axiom 8 holds if and only if the choice probabilities have the form as in (3.3) (suitably extended to the present context) with

(i) $$h(x) = \beta x + \kappa \quad and \quad u(j,w) = \frac{b_j\left(w^{\rho} - 1\right)}{\rho} + c_j$$

or

(ii) $$h(x) = \frac{\beta\left(x^{\theta} - 1\right)}{\theta} + \kappa \quad and \quad u(j,w) = b_j w^{\rho}$$

for $w \geq 0$*, where* $\beta > 0$*,* $b_j > 0, \kappa$ *and* c_j *are constants. In case (i)* c_j *may differ from* b_j / ρ *while in case (ii)* $\rho > 0$.

The proof of Theorem 7 is given in the Appendix.

Axiom 9

Let $\boldsymbol{g}_1, \boldsymbol{g}_2 \in S$*. If (3.3) holds, then*

$$G\left(h\left(V_1\left(\boldsymbol{g}_1\right)\right) - h\left(V_1\left(\boldsymbol{g}_2\right)\right)\right) = G\left(h\left(V_{\lambda}\left(\boldsymbol{g}_1\right)\right) - h\left(V_{\lambda}\left(\boldsymbol{g}_2\right)\right)\right)$$

for any real number $\lambda > 0$.

Axiom 9 is stronger than Axiom 8 because it postulates that the choice probabilities are invariant under scale transformations of the rewards.

Observe that Axioms 8 and 9 differ from the other axioms in that they are not "empirical", i.e., they do not have direct empirical nonparametric counterparts.

Corollary 3

Axiom 9 holds if and only if the choice probabilities have the form as in Theorem 4 with

(5.2) $$h(x) = \beta \log x + \kappa$$

for $u > 0$ *where* $\beta > 0$ *and* κ *are constants.*

Proof:

Clearly Axiom 9 implies Axiom 8. By Theorem 7 it therefore follows that either (i) or (ii) of Theorem 7 must hold. Evidently (i) cannot hold under Axiom 9. Consider next (ii). In this case Axiom 9 yields

$$\beta \cdot \frac{(\lambda x_1)^\theta - (\lambda x_2)^\theta}{\theta} = \beta \cdot \frac{\left(x_1^\theta - x_2^\theta\right)}{\theta}$$

for all $\lambda > 0$, where

$$x_j = \sum_{(k,w)\in X\times W} u(k,w) g_j(k,w)$$

for $j = 1,2$. The relation above implies that $\theta = 0$, in which case h reduces to (5.3).

Q.E.D.

Note that when $\beta = 1$, the choice model in Corollary 3 reduces to the so-called *Strict Expected Utility* model for uncertain outcomes proposed by Luce and Suppes (1965).

There are two alternative interpretations of the Axioms 1 to 3 and 6, 8 and 9 which represent extensions of the corresponding von Neumann-Morgenstern axioms. The first interpretation goes as follows: Consider an agent that participates in a large number of replications of a choice experiment. He may be boundedly rational in the sense that he has difficulties with assessing the precise value (to him) of the strategies in each single replication. This may be so even if he has no problem with assessing the values of the outcomes, simply because the evaluations of the respective lottery strategies do not follow immediately from the values of the outcomes and the outcome probabilities. The axioms state that while the agent is allowed to make “errors” when selecting strategies in each replication of the experiment (in the sense that his behavior is not consistent with the von Neumann-Morgenstern theory) he will still—in the aggregate sense specified in the axioms—behave according to the respective versions of the probabilistic extension of the von Neumann-Morgenstern theory.

In the alternative interpretation we consider a large observationally homogenous population. In this setting each agent in the population face the same choice experiment. While the behavior of each individual agent may be inconsistent with the von Neumann-Morgenstern theory the axioms above state that aggregate behavior in the population will be consistent with the probabilistic version of the theory.

Figure 1. Overview of axioms

Axiom 1
(i) If $P(\mathbf{g}_1,\mathbf{g}_2) \geq \frac{1}{2}$ and $P(\mathbf{g}_2,\mathbf{g}_3) \geq \frac{1}{2} \Rightarrow P(\mathbf{g}_1,\mathbf{g}_3) \geq \frac{1}{2}$
(ii) $P(\mathbf{g}_1,\mathbf{g}_2) + P(\mathbf{g}_2,\mathbf{g}_1) = 1$

Axiom 2
If $P(\mathbf{g}_1,\mathbf{g}_2) > \frac{1}{2}$ and $P(\mathbf{g}_2,\mathbf{g}_3) > \frac{1}{2}$,
there exists $\alpha,\beta \in (0,1)$ such that
$P(\alpha\mathbf{g}_1 + (1-\alpha)\mathbf{g}_3,\mathbf{g}_2) > \frac{1}{2}$ and $P(\mathbf{g}_2,\beta\mathbf{g}_1 + (1-\beta)\mathbf{g}_3) > \frac{1}{2}$

Axiom 3
$P(\mathbf{g}_1,\mathbf{g}_2) > \frac{1}{2}$
$\Downarrow$
$P(\alpha\mathbf{g}_1 + (1-\alpha)\mathbf{g}_3,\alpha\mathbf{g}_2 + (1-\alpha)\mathbf{g}_3) > \frac{1}{2}$
for all $\alpha \in [0,1]$

Axiom 4
$P(\mathbf{g}_1,\mathbf{g}_2) \geq \frac{1}{2} \Leftrightarrow P(\mathbf{g}_1,\mathbf{g}_3) \geq P(\mathbf{g}_2,\mathbf{g}_3)$

Axiom 5
(i) $P(\mathbf{g}_1,\mathbf{g}_2) \geq P(\mathbf{g}_3,\mathbf{g}_4) \Leftrightarrow P(\mathbf{g}_1,\mathbf{g}_3) \geq P(\mathbf{g}_2,\mathbf{g}_4)$
(ii) For y such that $P(\mathbf{g}_1,\mathbf{g}_3) \geq y \geq P(\mathbf{g}_1,\mathbf{g}_2)$, there is a $\mathbf{g} \in S$ such that $P(\mathbf{g}_1,\mathbf{g}) = y$
(iii) $P(\mathbf{g}_1,\mathbf{g}_2) + P(\mathbf{g}_2,\mathbf{g}_1) = 1$

Figure 1 (cont). Overview of axioms

Axiom 6

$$P(\mathbf{g}_1,\mathbf{g}_2) \geq P(\mathbf{g}_1^*,\mathbf{g}_2^*)$$

$$\Downarrow$$

$$P(\alpha\mathbf{g}_1+(1-\alpha)\mathbf{g}_3,\alpha\mathbf{g}_2+(1-\alpha)\mathbf{g}_3) \geq P(\alpha\mathbf{g}_1^*+(1-\alpha)\mathbf{g}_3,\alpha\mathbf{g}_2^*+(1-\alpha)\mathbf{g}_3)$$

for all $\alpha \in [0,1]$

Axiom 7 (IIA)

For given $\mathbf{g}_s \in S$, $P(\mathbf{g}_s,\mathbf{g}_r) \in (0,1)$ for all $\mathbf{g}_r \in S$,

$$P_B(\mathbf{g}_s) = P_A(\mathbf{g}_s)P_B(A),\ \mathbf{g}_s \in A \subset B,\ A,B \in \mathfrak{B}$$

Axiom 8

$$G\{h(V_1(\mathbf{g}_1)) - h(V_1(\mathbf{g}_2))\} \leq G\{h(V_1(\mathbf{g}_1^*)) - h(V_1(\mathbf{g}_2^*))\}$$

$$\Updownarrow$$

$$G\{h(V_\lambda(\mathbf{g}_1)) - h(V_\lambda(\mathbf{g}_2))\} \leq G\{h(V_\lambda(\mathbf{g}_1^*)) - h(V_\lambda(\mathbf{g}_2^*))\} \text{ for all } \lambda > 0,$$

where $V_\lambda(\mathbf{g}_s) = \sum_{(k,w)\in X\times W} u(k,\lambda w)g_s(k,w)$

Axiom 9

$$G\{h(V_1(\mathbf{g}_1)) - h(V_1(\mathbf{g}_2))\} = G\{h(V_\lambda(\mathbf{g}_1^*)) - h(V_\lambda(\mathbf{g}_2^*))\}$$

for all $\lambda > 0$

Axiom 5	$\Rightarrow$	**Axiom 1** **Axiom 4**
Axiom 6	$\Rightarrow$	**Axiom 3**
Axiom 7	$\Rightarrow$	**Axiom 4** **Axiom 5**
Axiom 8	$\Leftarrow$	**Axiom 9**

Figure 2. Relationship between axioms and binary choice probabilities

Axioms 1, 2, 3	$\Leftrightarrow$	**Corollary 1.** $\mathrm{P}(\mathbf{g}_1,\mathbf{g}_2) \geq \frac{1}{2} \Leftrightarrow \mathrm{V}(\mathbf{g}_1) \geq \mathrm{V}(\mathbf{g}_2)$ where $\mathrm{V}(\mathbf{g}_s) = \sum_{k \in X} u(k) g_s(k)$
Axioms 2, 3, 4	$\Leftrightarrow$	**Theorem 2.** $\mathrm{P}(\mathbf{g}_1,\mathbf{g}_2) = \mathrm{F}(\mathrm{V}(\mathbf{g}_1),\mathrm{V}(\mathbf{g}_2))$ for F strictly increasing in its first argument and strictly decreasing in the second
Axiom 5	$\Leftrightarrow$	**Theorem 3.** $\mathrm{P}(\mathbf{g}_1,\mathbf{g}_2) = \mathrm{G}(\mathrm{f}(\mathbf{g}_1) - \mathrm{f}(\mathbf{g}_2))$, for some function f that is unique up to a positive linear transformation and a c.d.f. G that is strictly increasing and continuous
Axioms 2, 3, 5	$\Leftrightarrow$	**Theorem 4.** $\mathrm{P}(\mathbf{g}_1,\mathbf{g}_2) = \mathrm{G}\{\mathrm{h}(\mathrm{V}(\mathbf{g}_1)) - \mathrm{h}(\mathrm{V}(\mathbf{g}_2))\}$ with h and G strictly increasing and G continuous
Axioms 2, 5, 6	$\Leftrightarrow$	**Theorem 5.** $\mathrm{P}(\mathbf{g}_1,\mathbf{g}_2) = \mathrm{G}\{\mathrm{V}(\mathbf{g}_1) - \mathrm{V}(\mathbf{g}_2)\}$

Figure 3. Relationship between axioms and multinomial choice probabilities

Axiom 7	⇔	$P_B(\mathbf{g}_s) = \frac{\exp(f(\mathbf{g}_s))}{\sum_{\mathbf{g}_r \in B} \exp(f(\mathbf{g}_r))}$, for some f
Axioms 2, 3, 7	⇔	**Theorem 6.** $P_B(\mathbf{g}_s) = \frac{\exp(h(V(\mathbf{g}_s)))}{\sum_{\mathbf{g}_r \in B} \exp(h(V(\mathbf{g}_r)))}$
Axioms 2, 6, 7	⇔	**Corollary 2.** $P_B(\mathbf{g}_s) = \frac{\exp(V(\mathbf{g}_s))}{\sum_{\mathbf{g}_r \in B} \exp(V(\mathbf{g}_r))}$

Relationship between axioms and binary choice probabilities for the case with monetary rewards

Axiom 8	⇔	**Theorem 7.** $h(x) = \beta x + \kappa$ and $u(j,w) = \frac{b_j(w^\rho - 1)}{\rho} + c_j$ or $h(x) = \frac{\beta(x^\theta - 1)}{\theta} + \kappa$ and $u(j,w) = b_j w^\rho$
Axiom 9	⇔	**Corollary 3.** $h(x) = \beta \log x + \kappa,\ \beta > 0$

Figures 1-3 display a convenient overview and summary of the results obtained in the paper. It is an important feature of Axioms 1 to 7 that they have direct empirical counterparts. Figure 2 emphasises the equivalences between sets of axioms and the structure of the respective choice probabilities. However, some of these choice probabilities depend on unknown functional forms (f, G, h). For example, all the binary choice probabilities depend on an unknown c.d.f. G. Only Theorem 5, Corollaries 2, 3, and Theorem 7 yield fully specified functional forms for the choice probabilities. As regards the results of Theorems 2 to 6 and Corollaries 1 and 2 the corresponding axioms can be used to test these models without relying on ad hoc functional form specifications. To carry out rigorous non-parametric tests of these axioms is in itself a complicated task. In fact, it seems that the general case with ordinal restrictions on choice probabilities of the type displayed in Figure 1 lies outside the scope of a large body of literature devoted to statistical hypotheses testing under ordinal constraints. As far as we know, only Iverson and Falmagne (1985) have explicitly addressed the challenge of developing test procedures for this type of setting. In particular, they discuss how one can test property (i) of Axiom 1 within a maximum likelihood setting.

6. A random utility representation

In this section we shall consider the problem of a random utility representation of the agent's preferences over lotteries that yield choice probabilities that satisfy the Axioms 2, 3 and 7. From the theory of discrete choice we know that the Luce choice model is consistent with an additive random utility representation in which the error terms are independent (across alternatives) extreme value c.d.f., $\exp\left(-e^{-x}\right)$. Let $\mathbf{g} \in S$ and let us for a moment assume there exists a random utility function U(**g**), $\mathbf{g} \in S$, that satisfies the axioms above. Here, the setting is not as simple as in the standard discrete choice case because S is *not* countable. Therefore, U(**g**) is a multiparameter stochastic process, i.e., a *random field*.

Theorem 8 (Random utility representation)

There exists a probability space and random variables $\left\{\varepsilon(\boldsymbol{g}), \boldsymbol{g} \in S\right\}$ *defined on it, such that* $\varepsilon(\boldsymbol{g}_s), s = 1,2,..., \boldsymbol{g}_s \in S$ *, are independent for distinct* $\boldsymbol{g}_1, \boldsymbol{g}_2,...,$ *and*

$$(6.1) \qquad P\left(\varepsilon(\boldsymbol{g}_s) \le y\right) = exp\left(-e^{-y}\right)$$

for $y \in R$ *. The random utility representation*

$$(6.2) \qquad U(\boldsymbol{g}) = h\left(V(\boldsymbol{g}_s)\right) + \varepsilon(\boldsymbol{g}),$$

for $\boldsymbol{g} \in S$ *, is consistent with Axioms 2, 3 and 7, i.e., for* $B \in \mathfrak{B}$

(6.3) $$P_B(\boldsymbol{g}_s) = P\left(U(\boldsymbol{g}_s) = \max_{\boldsymbol{g}_r \in B} U(\boldsymbol{g}_r)\right) = \frac{exp\left(h\left(V(\boldsymbol{g}_s)\right)\right)}{\sum_{\boldsymbol{g}_r \in B} exp\left(h\left(V(\boldsymbol{g}_r)\right)\right)}.$$

Proof:

By applying a special case of Kolmogorov's Theorem on the construction of random variables, the existence of the probability space on which the random field $\{\varepsilon(\mathbf{g}), \mathbf{g} \in S\}$ is defined follows. See for example the Corollary, page 18, in Lamperti (1966). This Corollary establishes the desired results for the case that is relevant in our context, namely when $\varepsilon(\mathbf{g}_s)$, $s = 1, 2, \ldots,$ are i.i.d. The choice probability in (4.3) follows from a well known result in discrete choice theory, see for example McFadden (1984). The result now follows from Theorem 6.

Q.E.D.

The other results obtained in Sections 4 and 5 follow as special cases of Theorem 8. It is also possible to give a random utility representation of the model in (6.2) by using the result of Corollary 1 in Dagsvik (2003). However, the choice model established in Theorem 2 is *not* consistent with a Random Utility Model.

Remark

The distribution function given in (6.1) is a so called type III *extreme value* distribution.[1] In statistics the extreme value distributions arise as the asymptotic distributions of the maximum of i.i.d. random variables. Many authors have studied this distribution in the context of the theory of discrete choice and random utility models, see for example McFadden (1973), Yellott (1977) and Strauss (1979). Under different regularity conditions they have demonstrated that (6.1) is the *only* distribution that implies a random utility representation that is consistent with Axiom 7.

7. An example

Consider the following choice setting. The agent has the choice of working in either of two wage work sectors or in a self-employment sector, denoted by alternative one, two and three, respectively. In wage work sector j he receives earnings w_j, $j = 1, 2$, with perfect certainty. In sector 3 earnings are uncertain. Hours of work in each sector are given. An example of a self-employment activity with fixed hours of work is found in the business of running a café or a bar with fixed opening hours. We

[1] There seems to be some confusion in the literature about the terminology. Some authors call (5.1) the type III extreme value distribution while other authors call it the type I extreme value distribution. Some authors also call it the Double Exponential Distribution.

assume that the agent has been running the business—or similar businesses—for many periods and consequently is able to calculate the empirical distribution of returns to his business. For simplicity we approximate this distribution with a discrete distribution. Let u(j,w) be the utility of working in sector j at wage income w. Let $g_3(w)$ be the lottery outcome probability that the agent receives wage w given that he chooses to work in the self-employment sector. The expected utilities of working in the wage sector reduce to $u(1, w_1)$ and $u(2, w_2)$, respectively, while the expected utility of working in sector 3 equals

$$\sum_{w \in W} u(3, w) g_3(w).$$

Under the assumptions of Corollary 2 it follows that the probability of working in wage sector j equals

$$(7.1) \qquad \tilde{P}_B(j) = \frac{\exp\left(u\left(j, w_j\right)\right)}{\sum_{s=1}^{2} \exp\left(u\left(s, w_s\right)\right) + \exp\left(\sum_{w \in W} u\left(3, w\right) g_3(w)\right)}$$

for $j = 1, 2$, where B = {1,2,3}. The probability of working in sector 3 equals

$$(7.2) \qquad \tilde{P}_B(3) = \frac{\exp\left(\sum_{w \in W} u\left(3, w\right) g_3(w)\right)}{\sum_{s=1}^{2} \exp\left(u\left(s, w_s\right)\right) + \exp\left(\sum_{w \in W} u\left(3, w\right) g_3(w)\right)}.$$

With convenient parametric specification of the utility function u(j,w) one can estimate the unknown parameters of the utility function by the method of maximum likelihood, provided data on agents' choices are available.

Alternatively, under the assumptions of Theorem 6 it follows that the probability of working in sector j becomes

$$(7.3) \qquad P_B(j) = \frac{\exp\left(h\left(u\left(j, w_j\right)\right)\right)}{\sum_{s=1}^{2} \exp\left(h\left(u\left(s, w_s\right)\right)\right) + \exp\left(h\left(\sum_{w \in W} u\left(3, w\right) g_3\left(w\right)\right)\right)}$$

for $j = 1, 2$, and similar expression for the probability of working in sector 3. Consequently, one can for example apply likelihood ratio tests procedures to test the hypothesis that h is linear against the alternative that h has the functional for given in Theorem 7. Recall that the maximum likelihood estimation procedure goes as follows: Let Y_j denote the number of agents that have chosen to work in sector j as observed in the data, and assume for simplicity that the choice probabilities above do not depend on observed individual characteristics. As is well known, the loglikelihood function can be expressed as

$$\text{LogL} = \sum_j Y_j \log P_B(j)$$

from which the unknown parameters are obtained by maximization of logL. The general case with individual characteristics is completely analogous. We refer to Amemiya (1985), and Ben-Akiva and Lerman (1985) for details about inference methods for discrete choice models.

8. Conclusion

In this paper we have developed a theory of probabilistic choice for risky choices based on different combinations of particular axioms. Some of the axioms are probabilistic extensions of the *Archimedean* and *Independence Axioms* in the von Neumann-Morgenstern theory of expected utility. We have explored the relationship between sets of axioms and the structure of the corresponding choice probabilities. In particular, one set of axioms implies a complete characterization of the functional form of the choice probabilities. Since most of the axioms proposed are non-parametric they can be utilized to carry out non-parametric tests of the respective structures of the choice probabilities. Finally, to illustrate the potential for applications we have discussed a concrete example.

References

Aczél, J. (1966): *Lectures on functional equations and their applications*. Academic Press, New York.

Amemiya, T. (1985): *Advanced Econometrics*. Basil Blackwell, Ltd., Oxford, UK.

Anderson, S.P., A. de Palma and J.F. Thisse (1992): *Discrete choice theory of product differentiation*. MIT Press, London.

Becker, G.M., M.H. De Groot and J. Marschak (1963a): Stochastic Models and Choice Behavior. *Behavioral Sciences*, **8**, 41-55.

Becker, G.M., M.H. De Groot and J. Marschak (1963b): An Experimental Study of Some Stochastic Models for Wagers. *Behavioral Sciences*, **8**, 199-202.

Ben-Akiva, M. and S.R. Lerman (1985): *Discrete choice analysis*. MIT Press, Cambridge.

Carbone, E. (1997): Investigation of Stochastic Preference Theory using Experimental Data. *Economic Letters*, **57**, 305-311.

Dagsvik, J.K. (2003): Nonparametric Identification of Random Utility Models. Mimeo, Statistics Norway.

Falmagne, J.C. (1985): *Elements of psychophysical theory*. Oxford University Press, London.

Fishburn, P.C. (1976): Binary Choice Probabilities between Gambles: Interlocking Expected Utility Models. *Journal of Mathematical Psychology*, **14**, 99-122.

Fishburn, P.C. (1978): A Probabilistic Expected Utility Theory of Risky Binary Choices. *International Economic Review*, **19**, 633-646.

Fishburn, P.C. (1998): Stochastic Utility. In S. Barberà, P.J. Hanmond and C. Seidl (eds.): *Handbook of Utility Theory*, Vol. 1; Principles. Kluwer Academic Publishers, London.

Harless, D.W. and C.F. Camerer (1994): The Predictive Utility of Generalized Expected Utility Theories. *Econometrica*, **62**, 1251-1290.

Hey, J.D. and C. Orme (1994): Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica*, **62**, 1291-1326.

Hey, J.D. (1995): Experimental Investigations of Errors in Decision Making under Risk. *European Economic Review*, **39**, 633-640.

Iverson, G. and J.C. Falmagne (1985): Statistical Issues of Measurement. *Mathematical Social Sciences*, **10**, 131-153.

Karni, E. and D. Schmeidler (1991): Utility Theory with Uncertainty. In W. Hildenbrand and H. Sonnenschein (eds.), *Handbook of Mathematical Economics*, Elsevier Science Publishers B.V., New York.

Loomes, G. and R. Sugden (1995): Incorporating a Stochastic Element into Decision Theories. *European Economic Review*, **39**, 641-648.

Loomes, G. and R. Sugden (1998): Testing Different Stochastic Specifications of Risky Choice. *Economica*, **65**, 581-598.

Luce, R.D. (1959): *Individual choice behavior: A theoretical analysis*. Wiley, New York.
Luce, R.D. and H. Raiffa (1957): *Games and decisions*. Wiley, New York.

Luce, R.D. and P. Suppes (1965): Preference, Utility, and Subjective Probability. In R.D. Luce, R.R. Bush, and E. Galanter (eds.). *Handbook of mathematical psychology, III*, 249-410. Wiley, New York.

McFadden, D. (1973): Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (ed.); *Frontiers in econometrics*, Academic Press.

McFadden, D. (1981): Econometric Models for Probabilistic Choice. In C. Manski and D. McFadden (eds.); *Structural analysis of discrete choice.* MIT Press, Cambridge, MA.

McFadden, D. (1984): Econometric Analysis of Qualitative Response Models. In Z. Griliches and M.D. Intriligator (eds.); *Handbook of econometrics*, Vol. II, Amsterdam.

Suppes, P., R.D. Luce, D. Krantz and A. Tversky (1989): *Foundations of measurement.* Vol. II. Academic Press, London.

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.

Strauss, D. (1979): Some Results on Random Utility Models. *Journal of Mathematical Psychology*, **17**, 35-52.

Thurstone, L.L. (1927a): A Law of Comparative Judgment. *Psychological Review*, **34**, 272-286.

Thurstone, L.L. (1927b): Psychological Analysis. *American Journal of Psychology*, **38**, 368-389.

Tversky, A. (1969): Intransitivity of Preferences. *Psychological Review*, **76**, 31-48.

Yellott, J.I., Jr. (1977): The Relationship between Luce’s Choice Axiom, Thurstone’s Theory of Comparative Judgment and the Double Exponential Distribution. *Journal of Mathematical Psychology*, **15**, 109-144.

Appendix

Proof of Theorem 2:

When the choice probabilities in Theorem 2 holds then evidently Axioms 2 and 4 hold. Consider next the "only if" part of the proof. Suppes et al. (1989), p. 412, have proved that Axiom 4 is equivalent to the representation

$$P(\mathbf{g}_1,\mathbf{g}_2)=\tilde{F}\left(f(\mathbf{g}_1),f(\mathbf{g}_2)\right)$$

for $\mathbf{g}_1,\mathbf{g}_2 \in S$, where f is an ordinal scale function defined on S and $\tilde{F}$ is a function that is strictly increasing in its first argument and strictly decreasing in the second. Hence

$$P(\mathbf{g}_1,\mathbf{g}_2)\geq\frac{1}{2} \Leftrightarrow f(\mathbf{g}_1)\geq f(\mathbf{g}_2)$$

and $\{f(\mathbf{g}),\mathbf{g}\in S\}$ therefore represents the binary relation $\succeq$ given in Definition 1. Accordingly $\succeq$ is a preference relation so that

$$f(\mathbf{g})=h\left(V(\mathbf{g})\right)$$

for some strictly increasing function h. Hence

$$P(\mathbf{g}_1,\mathbf{g}_2)=F\left(V(\mathbf{g}_1),V(\mathbf{g}_2)\right)$$

where

$$F(x,y)=\tilde{F}\left(h(x),h(y)\right).$$

Q.E.D.

Proof of Theorem 3:

Debreu (1958) has proved that Axiom 5 implies that there exists a cardinal representation $f(\mathbf{g}),\mathbf{g}\in S$, such that for $\mathbf{g}_1,\mathbf{g}_2,\mathbf{g}_3,\mathbf{g}_4\in S$

(A.1) $$P(\mathbf{g}_1,\mathbf{g}_2)\leq P(\mathbf{g}_3,\mathbf{g}_4)\Leftrightarrow f(\mathbf{g}_1)-f(\mathbf{g}_2)\leq f(\mathbf{g}_3)-f(\mathbf{g}_4),$$

where the inequality on one side is strict if and only if the inequality on the other side is strict. From (A.1) it follows that $\mathbf{g}_1$, $\mathbf{g}_2$, $\mathbf{g}_3$ and $\mathbf{g}_4$ satisfy $P(\mathbf{g}_1,\mathbf{g}_2)=P(\mathbf{g}_3,\mathbf{g}_4)$, if and only if $f(\mathbf{g}_1)-f(\mathbf{g}_2)=f(\mathbf{g}_3)-f(\mathbf{g}_4)$. But this means that we can write

$$P(\mathbf{g}_1,\mathbf{g}_2)=G\{f(\mathbf{g}_1)-f(\mathbf{g}_2)\}$$

for some suitable function G. Evidently, G(x) is strictly increasing and take values in [0,1]. Without loss of generality it can be chosen to be a cumulative distribution function. The Balance condition implies that $G(x)+G(-x)=1$, which means that G is symmetric. Recall that a cumulative distribution function is continuous to the right. Since G is symmetric it must also be continuous to the left. Hence G is continuous.

Next we shall prove uniqueness of G. Suppose (f_0, G_0) and (f_1, G_1) are two representations of the binary choice probabilities. Then

$$G_0\left(f_0(\mathbf{g}_1)-f_0(\mathbf{g}_2)\right)=G_1\left(f_1(\mathbf{g}_1)-f_1(\mathbf{g}_2)\right)$$

for any $\mathbf{g}_1, \mathbf{g}_2 \in S$. Since f_0 and f_1 are unique up to a linear transformation we can write

$$f_1(\mathbf{g}) = a f_0(\mathbf{g}) + b$$

for $\mathbf{g} \in S$, where a and b are constants and $a > 0$. This yields that

$$G_0\left(f_0(\mathbf{g}_1)-f_0(\mathbf{g}_2)\right)=G_1\left(a\left(f_0(\mathbf{g}_1)-f_0(\mathbf{g}_2)\right)\right)$$

which demonstrates that $G_0(x) = G_1(ax)$.

To prove that I is an interval, let $\mathbf{g}_0 \in S$ be a fixed point of reference. Let $\mathbf{g}_1, \mathbf{g}_2 \in S$ be such that $f(\mathbf{g}_2) \geq f(\mathbf{g}_1)$, and let $x \in \left[f(\mathbf{g}_1), f(\mathbf{g}_2)\right]$ be arbitrary. Hence, $f(\mathbf{g}_1)-f(\mathbf{g}_0) \leq x - f(\mathbf{g}_0) \leq f(\mathbf{g}_2)-f(\mathbf{g}_0)$, or equivalently

$$G^{-1}\left(P(\mathbf{g}_1, \mathbf{g}_0)\right) \leq x - f(\mathbf{g}_0) \leq G^{-1}\left(P(\mathbf{g}_2, \mathbf{g}_0)\right)$$

which yields

(A.2) $$P(\mathbf{g}_1, \mathbf{g}_0) \leq G\left(x - f(\mathbf{g}_0)\right) \leq P(\mathbf{g}_2, \mathbf{g}_0).$$

By Axiom 5(ii) there exists a $\mathbf{g}^* \in S$ such that $P\left(\mathbf{g}_0, \mathbf{g}^*\right) = G\left(f(\mathbf{g}_0) - x\right)$. Thus, (A.2) implies that

$$G\left(f(\mathbf{g}_0) - f(\mathbf{g}^*)\right) = P\left(\mathbf{g}_0, \mathbf{g}^*\right) = G\left(x - f(\mathbf{g}_0)\right)$$

so that $x = f\left(\mathbf{g}^*\right)$. Therefore, $x \in I$. Hence, we have proved that I is an interval.

Q.E.D.

Proof of Theorem 4:

When the choice probabilities in Theorem 4 holds then Axioms 2, 3 and 5 are satisfied. Consider the "only if" part. Debreu (1958) proved that Axiom 5 implies that there exists a mapping f from S to some interval such that for $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3, \mathbf{g}_4 \in S$

$$P(\mathbf{g}_1, \mathbf{g}_2) \geq P(\mathbf{g}_3, \mathbf{g}_4)$$

if and only if

$$f(\mathbf{g}_1) - f(\mathbf{g}_2) \geq f(\mathbf{g}_3) - f(\mathbf{g}_4).$$

Thus, with $\mathbf{g}_3 = \mathbf{g}_4$ we get

$$P(\mathbf{g}_1, \mathbf{g}_2) \geq 0.5 \Leftrightarrow f(\mathbf{g}_1) \geq f(\mathbf{g}_2)$$

and $\{f(\mathbf{g}), \mathbf{g} \in S\}$ therefore represents $\succeq$ on S. Consequently, $\succeq$ is a preference relation. But then, by Theorem 1, Axioms 2 and 3 imply that $f(\mathbf{g})$ must be a strictly increasing function h (say) of $V(\mathbf{g})$; that is

(A.3) $$f(\mathbf{g}) = h(V(\mathbf{g})).$$

Since Axiom 5 implies Theorem 3 we can combine (A.3) and (3.2) from which we get the desired result. Furthermore, by Theorem 3, V(·) is unique up to a linear transformation. Since evidently f(·) must also be unique up to a linear transformation we obtain the restrictions on $h(V(\cdot))$ stated in the theorem.

Q.E.D.

Proof of Theorem 5:

Note first that when choice probabilities of Theorem 5 holds it follows readily that Axioms 2, 5 and 6 are satisfied. Note next that when Axiom 6 holds, then if

(A.4) $$P(\mathbf{g}_1, \mathbf{g}_2) = P(\mathbf{g}_1^*, \mathbf{g}_2^*)$$

then

(A.5) $$P(\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_3, \alpha \mathbf{g}_2 + (1-\alpha)\mathbf{g}_3) = P(\alpha \mathbf{g}_1^* + (1-\alpha)\mathbf{g}_3, \alpha \mathbf{g}_2^* + (1-\alpha)\mathbf{g}_3)$$

for $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_1^*, \mathbf{g}_2^*, \mathbf{g}_3 \in S$ and $\alpha \in [0,1]$.

To realize this, note that

$$P(\mathbf{g}_1,\mathbf{g}_2)=P(\mathbf{g}_1^*,\mathbf{g}_2^*)$$

is equivalent to

$$P(\mathbf{g}_1,\mathbf{g}_2)\geq P(\mathbf{g}_1^*,\mathbf{g}_2^*) \quad \text{and} \quad P(\mathbf{g}_1,\mathbf{g}_2)\leq P(\mathbf{g}_1^*,\mathbf{g}_2^*).$$

When applying Axiom 6 twice with the inequality sign reversed the second time we therefore obtain (A.5).

Let $x_j = V(\mathbf{g}_j)$, $j=1,2,3$, where $V(\cdot)$ is given in Theorem 4. Then, since Axiom 6 implies Axiom 3, it follows that Theorem 4 holds. Accordingly, (3.3) yields

(A.6) $$P(\alpha\mathbf{g}_1+(1-\alpha)\mathbf{g}_3,\alpha\mathbf{g}_2+(1-\alpha)\mathbf{g}_3)=\tilde{G}\left(\frac{\tilde{h}(\alpha x_1+(1-\alpha)x_3)}{\tilde{h}(\alpha x_2+(1-\alpha)x_3)}\right)$$

where $\tilde{G}$ and $\tilde{h}$ are defined by $\tilde{G}(x)=G(e^x)$ and $\log\tilde{h}(x)=h(x)$. $h(\cdot)>0$ is a strictly increasing function defined on R.

Recall that h is strictly increasing. By (A.4), (A.5) and (A.6) we have that whenever x_j^* given by $x_j^*=V(\mathbf{g}_j^*)$, $\mathbf{g}_j^*\in S$, $j=1,2,$ satisfy

(A.7) $$\tilde{G}\left(\frac{\tilde{h}(x_1)}{\tilde{h}(x_2)}\right)=\tilde{G}\left(\frac{\tilde{h}(x_1^*)}{\tilde{h}(x_2^*)}\right)$$

then it follows that

(A.8) $$\tilde{G}\left(\frac{\tilde{h}(\alpha x_1+(1-\alpha)x_3)}{\tilde{h}(\alpha x_2+(1-\alpha)x_3)}\right)=\tilde{G}\left(\frac{\tilde{h}(\alpha x_1^*+(1-\alpha)x_3)}{\tilde{h}(\alpha x_2^*+(1-\alpha)x_3)}\right),$$

for any $\alpha\in[0,1]$. Without loss of generality we normalize V such that when $\mathbf{g}_0=(1,0,0,...)$, $V(\mathbf{g}_0)=0$. In particular, when $\mathbf{g}_3=\mathbf{g}_0$, $x_3=0$, it follows from (A.6) and (A.7) that whenever x_1^* and x_2^* are such that

(A.9) $$\frac{\tilde{h}(x_1)}{\tilde{h}(x_2)}=\frac{\tilde{h}(x_1^*)}{\tilde{h}(x_2^*)}$$

then

(A.10) $$\frac{\tilde{h}(\alpha x_1)}{\tilde{h}(\alpha x_2)}=\frac{\tilde{h}(\alpha x_1^*)}{\tilde{h}(\alpha x_2^*)}.$$

Note next that (A.9) and (A.10) imply that we can write

(A.11) $$\frac{\tilde{h}(\alpha x_1)}{\tilde{h}(\alpha x_2)}=f_\alpha\left(\frac{\tilde{h}(x_1)}{\tilde{h}(x_2)}\right)$$

for some strictly increasing continuous function f_α that depends on α. To realize this observe that

$$\frac{\tilde{h}(\alpha x_1)}{\tilde{h}(\alpha x_2)}$$

depends on x_1, x_2 solely through

$$\frac{\tilde{h}(x_1)}{\tilde{h}(x_2)}$$

due to the fact that the value of

$$\frac{\tilde{h}(\alpha x_1)}{\tilde{h}(\alpha x_2)}$$

is (by (A.10)) unchanged when (x_1, x_2) is replaced by (x_1^*, x_2^*) when (A.9) is satisfied.

Let $u=\tilde{h}(x_1)$, $1/v=\tilde{h}(x_2)$. From (A.11) we then get

(A.12) $$\frac{\tilde{h}\left(\alpha\tilde{h}^{-1}(u)\right)}{\tilde{h}\left(\alpha\tilde{h}^{-1}\left(\frac{1}{v}\right)\right)}=f_\alpha(uv).$$

From (A.12) it follows that $f_\alpha(u)$ is strictly increasing in u.

By letting u and v successively be equal to one, (A.12) implies that

(A.13) $$f_\alpha(u)=\frac{1}{f_\alpha\left(\frac{1}{u}\right)}.$$

Hence, by (A.12) and (A.13)

(A.14) $$f_\alpha(uv)=\frac{f_\alpha(u)}{f_\alpha\left(\frac{1}{v}\right)}=f_\alpha(u)f_\alpha(v).$$

Eq. (A.14) is a functional equation of the Cauchy type. Since $f_\alpha(u)$ is strictly increasing the only possible solution of (A.14) is given by

(A.15) $$f_\alpha(u) = u^{c(\alpha)}$$

where $c(\alpha)$ is a function of α, see for example Falmagne (1985), Theorem 3.4.

Recall that $\tilde{h}(\cdot)$ is only unique up to a multiplicative constant. Therefore, $\tilde{h}(\cdot)$ can be normalized such that $\tilde{h}(1) = 1$. From (A.11) and (A.15), with $x_1 = x$ and $x_2 = 1$, we obtain that

(A.16) $$h(\alpha x) = c(\alpha) h(x) + h(\alpha)$$

where

$$h(x) = \log \tilde{h}(x),$$

and h is defined on [0,1]. In the following it will be convenient to organize the rest of the proof in two cases.

Case 1. $c(\alpha)$ is a constant.

In this case (A.16) yields

$$h(\alpha x) = ch(x) + h(\alpha) = h(x\alpha) = ch(\alpha) + h(x)$$

and hence

$$(c-1)h(x) = (c-1)h(\alpha),$$

which must hold for all $x, \alpha \in [0,1]$. This implies that $c = 1$. Thus, (A.16) reduces to a well known Cauchy type functional equation which only continuous solution is

(A.17) $$h(x) = \beta \log x + \gamma$$

where β and γ are constants.

Case 2. $c(\alpha)$ is not a constant.

In this case there is at least one α, say α_0, such that $c(\alpha_0) \neq 1$. Hence, (A.16) leads to

(A.18) $$h(\alpha_0 x) = c(\alpha_0) h(x) + h(\alpha_0) = h(x\alpha_0) = c(x) h(\alpha_0) + h(x).$$

The last equation yields

(A.19) $$h(x)=\left(c(x)-1\right)b_0$$

where

$$b_0=\frac{h\left(\alpha_0\right)}{c\left(\alpha_0\right)-1}.$$

When (A.19) is inserted into (A.16) and terms are rearranged we obtain

(A.20) $$c\left(\alpha x\right)=c(\alpha)c(x)$$

for $\alpha, x\in\left[0,1\right]$. The only strictly increasing solution of (A.20) is

(A.21) $$c(\alpha)=\alpha^{\kappa}$$

for some constant κ (see Falmagne, Theorem 3.4). When (A.19) and (A.21) are combined we get

(A.22) $$h(x)=b_0\left(x^{\kappa}-1\right)$$

for $x\in\left[0,1\right]$. Note next that (A.7) and (A.8) imply that

(A.23) $$\left(\alpha x_1+\left(1-\alpha\right)x_3\right)^{\kappa}-\left(\alpha x_2+\left(1-\alpha\right)x_3\right)^{\kappa}=\left(\alpha x_1^*+\left(1-\alpha\right)x_3\right)^{\kappa}-\left(\alpha x_2^*+\left(1-\alpha\right)x_3\right)^{\kappa}$$

whenever

(A.24) $$x_1^{\kappa}-x_2^{\kappa}=\left(x_1^*\right)^{\kappa}-\left(x_2^*\right)^{\kappa}.$$

Now keep x_1^*, x_2^* and x_3 fixed and differentiate (A.23) with respect to x_1 subject to (A.24). This gives

(A.25) $$\left(\alpha x_1+\left(1-\alpha\right)x_3\right)^{\kappa-1}=\left(\alpha x_2+\left(1-\alpha\right)x_3\right)^{\kappa-1}\frac{dx_2}{dx_1}=\left(\alpha x_2+\left(1-\alpha\right)x_3\right)^{\kappa-1}\left(\frac{x_1}{x_2}\right)^{\kappa-1}.$$

Suppose that $\kappa\neq 1$. Then (A.25) implies that $x_1=x_2$, which is a contradiction. Also the solution given by (A.17) does not satisfy (A.25). We therefore conclude that $\kappa=1$, i.e.,

(A.26) $$h(x)=b_0\left(x-1\right).$$

Recall that the normalization $h(1)=0$ we adopted above was made purely for notational convenience so that the general form of h is $h(x)=b_0x+\gamma$ where γ is an arbitrary constant.

This completes the proof.

Q.E.D.

Proof of Theorem 6:

It follows immediately that the "if" part of the theorem is true. Consider the "only if" part. From the theory of discrete choice (see for example McFadden (1984)) it follows that Axiom 7 holds if and only if for any $B \in \mathfrak{B}$,

$$P_B(\mathbf{g}_s) = \frac{a(\mathbf{g}_s)}{\sum_{\mathbf{g}_r \in B} a(\mathbf{g}_r)}$$

where $a(\mathbf{g}_s), \mathbf{g}_s \in S$, is a positive scalar that depends solely on $\mathbf{g}_s$ and is unique apart from a multiplicative positive constant. Let $B = \{\mathbf{g}_r, \mathbf{g}_s\}$. Then

$$P(\mathbf{g}_s, \mathbf{g}_r) = \frac{a(\mathbf{g}_s)}{a(\mathbf{g}_s) + a(\mathbf{g}_r)} = \frac{1}{1 + a(\mathbf{g}_r)/a(\mathbf{g}_s)}.$$

Thus

$$P(\mathbf{g}_s, \mathbf{g}_r) \geq 0.5 \iff a(\mathbf{g}_s) \geq a(\mathbf{g}_r)$$

and $\{a(\mathbf{g}_s), \mathbf{g}_s \in S\}$ therefore represents $\succeq$ on S. Consequently, $\succeq$ is a preference relation. But then, by Theorem 1, $a(\mathbf{g}_s)$ must be a strictly increasing function h (say) of $V(\mathbf{g}_s)$. Hence, by Theorem 1

$$\log a(\mathbf{g}_s) = h(V(\mathbf{g}_s)).$$

Q.E.D.

Proof of Theorem 7:

Note first that it follows immediately that when (i) or (ii) in Theorem 7 holds then Axiom 8 is true. We shall next prove that (i), or (ii), also are necessary for Axiom 8 to be true. Without loss of generality we consider lotteries with only two outcomes and let $W = \{0, w_1, w_2, w_1^*, w_2^*\}$, $g_1(1, w_1) > 0$, $g_1(2,0) > 0$, $g_2(1, w_2) > 0$, $g_2(2,0) > 0$, $g_1^*(1, w_1^*) > 0$, $g_1^*(2,0) > 0$, $g_2^*(1, w_2^*) > 0$ and $g_2^*(2,0) > 0$, while the remaining outcome probabilities on $X \times W$ are equal to zero. Hence

$$V_\lambda(\mathbf{g}_s) = \tilde{V}_\lambda(\mathbf{g}_s, w_s) \equiv g_s(1, w_s)\big(u(1, \lambda w_s) - u(1,0)\big) + u(2,0)$$

and

$$V_\lambda(\mathbf{g}_s^*) = \tilde{V}_\lambda(\mathbf{g}_s^*, w_s^*) \equiv g_s^*(1, w_s^*)\big(u(1, \lambda w_s^*) - u(2,0)\big) + u(2,0)$$

for $s=1,2$. Let $\tilde{G}$ and $\tilde{h}$ be defined by $\tilde{G}(x)=G\left(e^x\right)$ and $\ln\tilde{h}(x)=h(x)$. Without loss of generality we let $g_s\left(j,w_j\right)=g_s(j)$ and $g_s^*\left(j,w_j^*\right)=g_s^*(j),\ j=1,2$, be independent of the values of w_j and w_j^*, respectively. By Axiom 8, if

$$\tilde{G}\left(\tilde{h}\left(\tilde{V}_1\left(\mathbf{g}_1,w_1\right)\right)/\tilde{h}\left(\tilde{V}_1\left(\mathbf{g}_2,w_2\right)\right)\right)\le\tilde{G}\left(\tilde{h}\left(\tilde{V}_1\left(\mathbf{g}_1^*,w_1^*\right)\right)/\tilde{h}\left(\tilde{V}_1\left(\mathbf{g}_2^*,w_2^*\right)\right)\right),$$

then for all $\lambda>0$

$$\tilde{G}\left(\tilde{h}\left(\tilde{V}_\lambda\left(\mathbf{g}_1,w_1\right)\right)/\tilde{h}\left(\tilde{V}_\lambda\left(\mathbf{g}_2,w_2\right)\right)\right)\le\tilde{G}\left(\tilde{h}\left(\tilde{V}_\lambda\left(\mathbf{g}_1^*,w_1^*\right)\right)/\tilde{h}\left(\tilde{V}_\lambda\left(\mathbf{g}_2^*,w_2^*\right)\right)\right).$$

We can now apply Theorem 14.19, p.338, in Falmagne (1985), which yields

$$\begin{aligned}&\tilde{G}\left(\tilde{h}\left(\tilde{V}_1\left(\mathbf{g}_1,w_1\right)\right)/\tilde{h}\left(\tilde{V}_1\left(\mathbf{g}_2,w_2\right)\right)\right)\equiv G\left(h\left(\tilde{V}_1\left(\mathbf{g}_1,w_1\right)\right)-h\left(\tilde{V}_1\left(\mathbf{g}_2,w_2\right)\right)\right)\\ &=F^*\left(\frac{b_1\left(w_1^\rho-1\right)-b_2\left(w_2^\rho-1\right)}{\rho}\right)\end{aligned}\tag{A.27}$$

for some strictly increasing continuous function F^*, where b_1, b_2 and ρ are constants. As usual, we define

$$\frac{w^\rho-1}{\rho}=\log w$$

when $\rho=0$. Let $M=G^{-1}F^*$. Hence (A.27) yields

$$h\left(\tilde{V}_1\left(\mathbf{g}_1,w_1\right)\right)-h\left(\tilde{V}_1\left(\mathbf{g}_2,w_2\right)\right)=M\left(\frac{b_1\left(w_1^\rho-1\right)-b_2\left(w_2^\rho-1\right)}{\rho}\right).\tag{A.28}$$

With $w_2=1$, (A.28) gives

$$h\left(\tilde{V}_1\left(\mathbf{g}_1,w_1\right)\right)=h\left(\tilde{V}_1\left(\mathbf{g}_2,1\right)\right)+M\left(\frac{b_1\left(w_1^\rho-1\right)}{\rho}\right).\tag{A.29}$$

Similarly, $w_1=1$ gives

$$h\left(\tilde{V}_1\left(\mathbf{g}_2,w_2\right)\right)=h\left(\tilde{V}_1\left(\mathbf{g}_1,1\right)\right)-M\left(\frac{-b_2\left(w_2^\rho-1\right)}{\rho}\right).\tag{A.30}$$

Let $\mathbf{g}_1=\mathbf{g}_2$ and

$$x_j = \frac{b_j\left(w_j^\rho - 1\right)}{\rho}.$$

It follows that (A.28), (A.29) and (A.30) imply that

(A.31) $$M\left(x_1 - x_2\right) = M\left(x_1\right) + M\left(-x_2\right) + h\left(V_1\left(\mathbf{g}_2\right)\right) - h\left(V_1\left(\mathbf{g}_1\right)\right) = M\left(x_1\right) + M\left(-x_2\right).$$

Hence (A.31) implies that

(A.32) $$M(x) + M(y) = M\left(x + y\right)$$

for x and y that belong to a suitable interval. Since M is continuous we must have that

(A.33) $$M(x) = bx$$

where b is a constant (see for example Falmange, 1985, Theorem 3.2). Consequently, we have established that for $g\left(j, w\right) = g(j)$ we have

(A.34) $$h\left(\tilde{V}_1\left(\mathbf{g}, w\right)\right) = h\left(g(1)\left(u\left(1, w\right) - u\left(2, 0\right)\right)\right) = b\left(\frac{w^\rho - 1}{\rho}\right) + c$$

where b and c may depend on **g**. Define z, f(·) and $\psi(\cdot)$ by

$z = u(1, w) - u(2, 0), f\left(g(1)z\right) = h\left(g(1)z + u(2, 0)\right)$ and

(A.35) $$\psi(z) = \frac{w^\rho - 1}{\rho}.$$

The function ψ is well defined and continuous because u(1,w) is strictly increasing and continuous. Hence we realize that (A.34) has a structure that is equivalent to the functional equation

(A.36) $$f\left(gz\right) = c\left(g(1)\right) + b\left(g(1)\right)\psi(z).$$

The solution of (A.36) can be found in Falmagne (1985), see case (iv) in Table 3.10, p. 89. The solution is given by

(A.37) $$f(z) = \beta_0\left(1 - z^\theta\right) + \beta_1 + \beta_2,$$

(A.38) $$c\left(g(1)\right) = \left(\beta_0 + \beta_1\right)\left(1 - g(1)^\theta\right) + \beta_2,$$

(A.39) $$\psi(z) = \frac{\beta_0}{\beta_3}\left(1 - z^\theta\right) + \frac{\beta_1}{\beta_3}$$

and

(A.40) $$b\left(g(1)\right) = \beta_3 g(1)^{\theta}$$

where β_0, β_1, β_2, β_3 and θ are constants. Hence, we can write

(A.41) $$f(z) = h\left(z + u(2,0)\right) = \alpha + \gamma z^{\theta}$$

for suitable constants α and γ. If $u(2,0) = 0$ then clearly $h(z) = \alpha + \gamma z^{\theta}$ with α and γ independent of u(2,0). If $u(2,0) \neq 0$ then evidently θ must be equal to one and γ must be independent of u(2,0) while $\alpha = \gamma u(2,0)$. Furthermore, it follows from (A.39), together with the requirement that $\psi(1) = 0$, that u(1,w) has the form

(A.42) $$u(1,w) = \tilde{b}\left(\frac{w^{\kappa} - 1}{\kappa}\right) + \tilde{c}$$

where $\tilde{b} > 0, \tilde{c}$ and κ are suitable constants.

Hence the proof in complete.

Q.E.D.