Stochastic Expected Utility for Binary Choice: A 'Modular' Axiomatic Foundation

Matthew Ryan*

This version: March 2019

Abstract

We present new axiomatisations for various models of binary stochastic choice that may be characterised as "expected utility maximisation with noise". These include axiomatisations of strictly (Ryan 2018a) and simply (Tversky and Russo, 1969) scalable models, plus strict (Ryan, 2015) and strong (Debreu, 1958) Fechner models. Our axiomatisations complement the important contributions of Blavatskyy (2008) and Dagsvik (2008). Our representation theorems set all models on a common axiomatic foundation, progressively augmented by additional axioms necessary to characterise successively more restrictive models. The key is a decomposition of Blavatskyy's (2008) *common consequence independence* axiom into two parts: one that underwrites the linearity of utility and another than underwrites the Fechnerian structure of noise. This decomposition is also useful for testing Fechnerian models, as we discuss.

JEL codes: D01, D81

Keywords: stochastic choice, expected utility, scalability, Fechner

*My thanks to Aurélien Baillon for several useful suggestions which are reflected in the present version of the paper. I have also benefitted from the comments of audiences at Queen Mary University of London, the 37th Australasian Economic Theory Workshop (University of Technology Sydney) and the DECIDE Workshop on Experimental Economics (University of Auckland). Previous drafts have circulated under the title "Stochastic Expected Utility for Binary Choice: New Representations". Address for correspondence: School of Economics, Auckland University of Technology, PB 92006, Auckland 1142, New Zealand. Email: `mryan@aut.ac.nz`

1 Introduction

Since John Hey issued his provocative challenge in the mid-1990s, there has emerged a sizeable revisionist literature on the descriptive merits of expected utility (EU). Hey hypothesised that "one can explain experimental analyses of decision making under risk better (and simpler) as EU plus noise – rather than through some higher level functional – as long as one specifies the noise appropriately" (Hey, 1995, p.640). Testing this hypothesis means viewing the evidence through the lens of a *probabilistic* model of choice. Such models specify a utility function plus an auxiliary function that converts the utilities of the available alternatives into choice probabilities; that is, a model of the preference "signal" as well as the behavioural "noise".

This revisionist literature may be broadly separated into two categories. In the first are papers that test necessary conditions for specific models of probabilistic choice, to see which may be rejected. A notable example is the paper by Loomes and Sugden (1998), who test (*inter alia*) EU embedded in a Fechnerian noise structure. This structure implies that the probability of choosing lottery α over lottery β (in a binary choice) is a non-decreasing function of the difference between the expected utility of α and the expected utility of β. Loomes and Sugden test a necessary consequence of this model, which Blavatskyy (2008) calls *common consequence independence (CCI).* Their experimental data firmly reject CCI, and hence the "EU plus Fechnerian noise" model.

The second category comprises studies that directly compare various combinations of utility function and noise structure to identify the combination that achieves the best fit to the data, after applying an appropriate statistical penalty for (lack of) parsimony. Examples include Buschena and Zilberman (2000), Blavatskyy and Pogrebna (2010), and Conte, Hey and Moffatt (2011).

Hey himself favours the second of these two approaches – see Hey (2014). A major weakness of the first approach is that model rejections rarely provide useful guidance on how to improve the model; they don't indicate which aspect of the model is at fault or how to repair the faulty component. The necessary conditions tested in these studies typically embody *joint* hypotheses about both the utility structure and the noise. Loomes and Sugden's rejection of CCI, for example, is a rejection of the joint hypothesis that utility is mixture-linear and noise Fechnerian. Should we blame Fechner (as Hey conjectures) or von Neumann and Morgenstern (as Loomes and Sugden are inclined to do), or are both aspects of the model at fault?

In this paper we present some theoretical results that may assist experimentalists to answer such questions. Theorem 3.3 decomposes the CCI condition into two parts – one that ensures linear utility and another that underwrites a Fechnerian noise structure *conditional on utility being mixture-linear* (our Axioms 5 and 6 respectively). This facilitates more discriminating tests of the model: if CCI is rejected but the weaker condition underwriting Fechnerian noise (our Axiom 6) is not, then the finger of blame points firmly at von Neumann and Morgenstern.

This result is obtained as part of a broader theoretical exercise. We develop new axiomatisations of models that embed EU in two types of Fechnerian noise structures: one in which binary choice probabilities are expressed as a *non-decreasing* function of EU differences, and a stronger type in which the function that converts EU differences into choice probabilities is *strictly increasing* (see Definitions 2 and 4). These representation theorems (Theorems 3.3 and 3.5 below) complement the important earlier work of Blavatskyy (2008), who provided an axiomatisation of the weaker model, and Dagsvik (2008), who axiomatised a version of the stronger one in which choice probabilities are also required to vary *continuously* with EU differences. We provide new axiomatisations for both of these models (see Theorems 3.3 and 3.9).

Unlike those of Blavatskyy (2008) and Dagsvik (2008), our axiomatisations have a modular structure that separates the conditions guaranteeing linear utility from those required for Fechnerian noise. Along the way we obtain axiomatisations of binary choice probabilities that are *scalable* with respect to a mixture-linear utility function. These models generalise their Fechnerian counterparts by relaxing the requirement that choice probabilities depend only on utility *differences* – see Definitions 1 and 3.

Like Ryan (2018a), this paper contributes to the literature on what has been called *Luce's challenge* (Regenwetter, Dana and Davis-Stober, 2011): the challenge of identifying methods for converting axiomatic characterisations of preferences with a representation within a particular class of utility functions into axiomatic characterisations of choice probabilities that embody the noisy maximisation of some utility function within the given class.

Our new axiomatisations have one further ancillary benefit. Despite the close similarity between their respective Fechnerian models, the axiomatisations offered by Blavatskyy (2008) and Dagsvik (2008) are very different, as are the proof strategies. Our results set both models upon a common axiomatic foundation and harmonise the proofs of the respective representation theorems. We show that Dagsvik's model can be characterised by making two modifications to the axioms of Blavatskyy: strengthening the *strong stochastic transitivity (SST)* axiom (compare Definitions 2 and 7 below) and adding Debreu's (1958) *solvability* condition (Theorem 3.9). The strengthening of SST is needed to obtain the strict monotonicity of the function that converts EU differences into choice probabilities, while solvability adds the required continuity. This makes the model differences axiomatically transparent.

The following section reviews some basic ideas from binary stochastic choice. Our representation results are given in Section 3. Some implications for experimental testing are briefly discussed in Section 4. The Appendix contains proofs of all results.

2 Binary choice probabilities

Let A be a *mixture set* of alternatives,[1] such as the set of lotteries studied by Blavatskyy (2008) and Dagsvik (2008). If $a, b \in A$ and $\lambda \in [0,1]$ we use $a\lambda b$ to denote the λ-mixture of a and b. In particular, $a1b = a$ and $a0b = b$. A function $u : A \to \mathbb{R}$ is *mixture-linear* if $u(a\lambda b) = \lambda u(a) + (1-\lambda) u(b)$ for any $a, b \in A$ and any $\lambda \in [0,1]$. If u is mixture-linear then $u(A)$ is an interval (Ryan, 2018a, Lemma 2).

The objects of analysis will be functions $P : A^2 \to [0,1]$. If $a \neq b$ then $P(a,b)$ is interpreted as the probability with which a given decision-maker chooses alternative a from the binary choice set $\{a, b\}$. If $a = b$ then no behavioural interpretation is given. We call such a function a *binary choice probability (BCP)*. It is natural to require that any BCP satisfy $P(a,b) + P(b,a) = 1$ for any $a, b \in A$ (hence $P(a,a) = \frac{1}{2}$ for any $a \in A$, which fixes the non-behavioural terms), but we follow Blavatskyy (2008) and Dagsvik (2008) in imposing this as an axiomatic restriction rather than as part of the definition of a BCP; it is the *balance* axiom (Axiom 1) below.

Associated with a binary choice probability, P, is the following binary relation on A:

$$a \succsim^P b \quad \Leftrightarrow \quad P(a,b) \geq \frac{1}{2} \tag{1}$$

That is, $a \succsim^P b$ iff the decision-maker is at least as likely to choose a as to choose b in a binary choice. The asymmetric and symmetric parts of $\succsim^P$, denoted $\succ^P$ and $\sim^P$ respectively, are defined in the usual way. We will refer to $\succsim^P$ as the *base relation* for P, by analogy with the theory of deterministic choice functions.[2] Note that the base relation is complete by construction but is not transitive unless P satisfies the following condition, known as *weak stochastic transitivity (WST)*:

$$\min\{P(a,b),\ P(b,c)\} \ \geq \ \frac{1}{2} \quad \Rightarrow \quad P(a,c) \ \geq \ \frac{1}{2}$$

for all $a, b, c \in A$. If the function $u : A \to \mathbb{R}$ represents the base relation (that is: $a \succsim^P b$ iff $u(a) \geq u(b)$ for any $a, b \in A$) then we call u a *weak utility* for P (Marschak, 1960).

3 Models and representations

In this section we recall a range of "single utility" models for BCPs, based on scalability or Fechnerian structures, and provide axiomatisations for the special case of each model in which utility is required to be mixture-linear.

[1] See Fishburn (1982, Section 2.1) for a formal definition. For readers who are unfamilar with mixture sets, there is some loss of generality but no damage to comprehension from treating A as a convex subset of Euclidean space.

[2] The reader is warned that this terminology is not standard for stochastic choice.

3.1 Models with weak monotonicity

The following two classes of models were defined in Ryan (2018b) and Ryan (2015) respectively:

Definition 1 *A binary choice probability P is* ***strictly scalable*** *iff there exists a weak utility function $u : A \to \mathbb{R}$ for P, and a function $F : u(A) \times u(A) \to [0,1]$ that is non-decreasing (respectively, non-increasing) in its first (respectively, second) argument and satisfies $F(x,y) + F(y,x) = 1$ for all $x, y \in u(A)$, such that*

$$P(a,b) = F(u(a), u(b))$$

for all $a, b \in A$. In this case, we say that P is strictly scalable by (u, F).

Definition 2 *A binary choice probability P has a* ***strict Fechner model*** *iff there exists a weak utility function $u : A \to \mathbb{R}$ for P, and a non-decreasing function $G : \Gamma \to [0,1]$, where $\Gamma = u(A) - u(A)$,[3] such that $G(x) + G(-x) = 1$ for all $x \in \Gamma$ and*

$$P(a,b) = G(u(a) - u(b))$$

for all $a, b \in A$. In this case, we say that (u, G) is a strict Fechner model for P.

Of course, any BCP with a strict Fechner model is strictly scalable: P has a strict Fechner model iff it is strictly scalable by some (u, F) in which $F(x,y) = F(x',y')$ whenever $x - y = x' - y'$. Note further that if P is strictly scalable then it satisfies $P(a,b) + P(b,a) = 1$.

Strict scalability captures the most basic sense in which a BCP might be said to describe a process of "noisy" utility maximisation. For a strict Fechner model, we impose the additional constraint that this noise depend only on the utility difference between the alternatives.

Strict scalability is a mild strengthening of the classical notion of *monotone scalability* (Fishburn, 1973), which differs from strict scalability only in not requiring u to be a weak utility for P. A utility function that does not represent $\succsim^P$ is somewhat awkward to interpret in the context of such a model, so it is natural to restrict attention to BCPs which are monotone scalable with respect to some weak utility for P; that is, to BCPs that are strictly scalable. As noted in Ryan (2015), this restriction is implicit in the analysis of Blavatskyy (2008).

[3] Recall that if $E \subseteq \mathbb{R}$ and $F \subseteq \mathbb{R}$ then $E - F$ denotes the set $\{x - y \mid x \in E \text{ and } y \in F\} \subseteq \mathbb{R}$.

Blavatskyy (2008) shows that the following axioms are necessary and sufficient for the existence of a strict Fechner model with a *mixture-linear* utility function:[4]

Axiom 1 (Balance) *For any* $a, b \in A$, $P(a, b) + P(b, a) = 1$.

Axiom 2 (Strong Stochastic Transitivity [SST]) *For any* $a, b, c \in A$,

$$\min\{P(a, b)\,,\ P(b, c)\} \ \geq\ \frac{1}{2} \quad \Rightarrow \quad P(a, c) \ \geq\ \max\{P(a, b)\,,\ P(b, c)\}\,.$$

Axiom 3 (Continuity) *For any* $a, b, c \in A$ *the following sets are closed*

$$\left\{\lambda \in [0, 1] \;\middle|\; P(a\lambda b, c) \geq \frac{1}{2}\right\}$$

$$\left\{\lambda \in [0, 1] \;\middle|\; P(a\lambda b, c) \leq \frac{1}{2}\right\}$$

Axiom 4 (Common Consequence Independence [CCI]) *For any* $a, b, c, d \in A$ *and any* $\lambda \in [0, 1]$, *we have* $P(a\lambda c, b\lambda c) = P(a\lambda d, b\lambda d)$.

Theorem 3.1 (Blavatskyy, 2008 [modified by Ryan, 2015]) *The binary choice probability P has a strict Fechner model iff it satisfies Axioms 1-4.*

As suggested in the Introduction, CCI does dual service in this representation result. Axioms 1-3 guarantee neither strict scalability with respect to a mixture-linear utility "scale" nor the existence of a strict Fechner model. This is confirmed by Example 3.1 below. Before analysing this example, we need one preliminary result (which may be of independent interest).

Lemma 3.1 *Let P be a BCP that is strictly scalable by some (u, F), with $u(A)$ a non-degenerate interval and F continuous in each argument. If (v, G) is a strict Fechner model for P then $v = h \circ u$ for some continuous and strictly increasing function $h : u(A) \to \mathbb{R}$.*

[4] Blavatskyy (2008) calls Axiom 1 "completeness" but we adopt Dagsvik's terminology here. Blavatskyy also included a fifth axiom, *interchangeability*, but this is implied by balance and strong stochastic transitivity – see Ryan (2015). Furthermore, while Blavatskyy proves his result for the special case in which A is the unit simplex in $\mathbb{R}^n$, we state it for an arbitrary mixture set. Our statement is therefore somewhat more general than Blavatskyy's. This generalised version of Blavatskyy's result is a corollary of Theorem 3.3 below.

Example 3.1 *Let A be the unit simplex in $\mathbb{R}^3$ (endowed with the usual mixture operation). Think of this as the set of all lotteries over some fixed set $X = \{x_1, x_2, x_3\}$ of "prizes". Figure 1 depicts A in the form of a Marschak-Machina (MM) triangle, with the probability of x_1 measured on the horizontal axis and the probability of x_3 on the vertical. If $b = (b_1, b_2, b_3) \in A$ then we abuse notation and also use b to denote the corresponding point (b_1, b_3) in the triangle. Let $\underline{a} = (1, 0, 0)$ and $\overline{a} = (0, 0, 1)$. The points $\underline{a}$ and $\overline{a}$ are indicated in Figure 1. (In the following, it will be useful to imagine that x_3 is the best prize and x_1 the worst.)*

Note that the line joining any point in the triangle to the point $(1, 1)$ outside the triangle has a unique intersection with the hypotenuse. For each $a \in A$, define $\lambda_a \in [0, 1]$ by the requirement that the line joining the point $a \in A$ to the point $(1, 1)$ passes through $\overline{a}\lambda_a\underline{a}$. Figure 1 illustrates. Now define P as follows:

$$P(a, b) = \begin{cases} \frac{1}{2} + \frac{1}{2}(\lambda_a)^2(\lambda_a - \lambda_b) & \text{if } \lambda_a \geq \lambda_b \\ \\ \frac{1}{2} - \frac{1}{2}(\lambda_b)^2(\lambda_b - \lambda_a) & \text{if } \lambda_a < \lambda_b \end{cases}$$

for each $a, b \in A$. It is easy to see that the range of P is contained in $[0, 1]$ so P is a binary choice probability.

Note that

$$P(a, b) \geq \frac{1}{2} \quad \Leftrightarrow \quad \lambda_a \geq \lambda_b$$

for any $a, b \in A$. Defining $\hat{u} : A \to \mathbb{R}$ by $\hat{u}(a) = \lambda_a$ and $F : [0, 1]^2 \to [0, 1]$ by

$$F(x, y) = \begin{cases} \frac{1}{2} + \frac{1}{2}(x)^2(x - y) & \text{if } x \geq y \\ \\ \frac{1}{2} - \frac{1}{2}(y)^2(y - x) & \text{if } x < y \end{cases}$$

we see that P is strictly scalable by $(\hat{u}, F)$. In particular, $\hat{u}$ is a weak utility for P with $\hat{u}(A) = [0, 1]$, and F is strictly increasing (respectively, strictly decreasing) in its first (respectively, second) argument and satisfies $F(x, y) + F(y, x) = 1$ for any $x, y \in [0, 1]$. It follows by Ryan (2018b, Lemma 11 and Theorem 14) that P satisfies Axioms 1-2. Since λ_a varies continuously with $a \in A$, and $F(x, y)$ is continuous in x for any y, we deduce that $P(a\lambda b, c)$ varies continuously with λ. Hence, Axiom 3 is satisfied. In summary: P is a BCP that satisfies Axioms 1-3, and any weak utility for P is a strictly increasing function of $\hat{u}$.

The contours (level sets) of any weak utility for P are therefore described by the lines emanating from the point $(1, 1)$ in Figure 1 (or rather, by the intersections of such lines with the triangle). It is obvious that no such utility function can satisfy mixture-independence. Thus, P is not strictly scalable with respect to any mixture-linear utility "scale".

It remains to show that P is has no strict Fechner model. Suppose it did. Then, using Lemma 3.1, there exists some strictly increasing and continuous function $h : [0,1] \to \mathbb{R}$ and some non-decreasing function $G : \hat{\Gamma} \to [0,1]$, where $\hat{\Gamma} = h([0,1]) - h([0,1])$, such that

$$P(a,b) = F(\lambda_a, \lambda_b) = G(h(\lambda_a) - h(\lambda_b))$$

for all $a, b \in A$. Since F is strictly increasing (respectively, strictly decreasing) in its first (respectively, second) argument, G must be strictly increasing on its domain.[5] *Thus:*

$$P(a,b) \geq P(c,d) \quad \Leftrightarrow \quad F(\lambda_a, \lambda_b) \geq F(\lambda_c, \lambda_d) \quad \Leftrightarrow \quad h(\lambda_a) - h(\lambda_b) \geq h(\lambda_c) - h(\lambda_d)$$

for any $a, b, c, d \in A$. From the equivalence of the first and last of these inequalities, it follows that P must satisfy:

$$P(a,b) = P(c,d) \quad \Leftrightarrow \quad P(a,c) = P(b,d) \tag{2}$$

for any $a, b, c, d \in A$. Let $a, b, c, d \in A$ be chosen such that $\lambda_a = 1$, $\lambda_b = \frac{7}{8}$, $\lambda_c = \frac{1}{2}$ and $\lambda_d = 0$. Then

$$\begin{aligned} (\lambda_a)^2 (\lambda_a - \lambda_b) = \frac{1}{8} = (\lambda_c)^2 (\lambda_c - \lambda_d) \quad &\Rightarrow \quad F(\lambda_a, \lambda_b) = F(\lambda_c, \lambda_d) \\ &\Leftrightarrow \quad P(a,b) = P(c,d) \end{aligned}$$

but

$$\begin{aligned} (\lambda_a)^2 (\lambda_a - \lambda_c) = \frac{1}{2} < \left(\frac{7}{8}\right)^3 = (\lambda_b)^2 (\lambda_b - \lambda_d) \quad &\Rightarrow \quad F(\lambda_a, \lambda_c) < F(\lambda_b, \lambda_d) \\ &\Leftrightarrow \quad P(a,c) < P(b,d) \end{aligned}$$

which contradicts (2).

[5] This fact is somewhat easier to believe than it is to show. To convince yourself it is true, let $\Sigma = h([0,1])$ and draw the following graph. Measure $h(\lambda_a)$ along the horizontal axis and $h(\lambda_b)$ along the vertical. Thus, noting that Σ is an interval by the continuity of h, the set $\Sigma \times \Sigma$ is a square bisected diagonally by the 45 degree line of the graph – see Figure 3 in the Appendix. Let $\hat{G}(x,y) = G(x-y)$ for any $(x,y) \in \Sigma \times \Sigma$. The function $\hat{G}$ is therefore constant along any line parallel to the 45 degree line (or rather, along the portion of such a line that intersects $\Sigma \times \Sigma$). The contours of $\hat{G}$ are also symmetric about the 45 degree line, since P satisfies balance (Axiom 1). It therefore suffices to consider the portion of $\Sigma \times \Sigma$ that lies on or below the 45 degree line, and to show that $\hat{G}$ is strictly increasing in $x - y$ on this part of its domain. Let $z \in \Sigma \times \Sigma$ and $z' \in \Sigma \times \Sigma$ be such that $z_1 \geq z_2$ and $z'_1 \geq z'_2$, with $z_1 - z_2 > z'_1 - z'_2$ (so z' lies closer to the 45 degree line than z). If $(z_1, -z_2) > (z'_1, -z'_2)$, then $F(h^{-1}(z_1), h^{-1}(z_2)) > F(h^{-1}(z'_1), h^{-1}(z'_2))$ and hence $\hat{G}(z) > \hat{G}(z')$. Otherwise, z must lie to the northeast or to the southwest of z'. In either case, we can move along the line through z' that is parallel to the 45 degree line to reach some point z'' that is due west of z and conclude that $\hat{G}(z) > \hat{G}(z'') = \hat{G}(z')$.

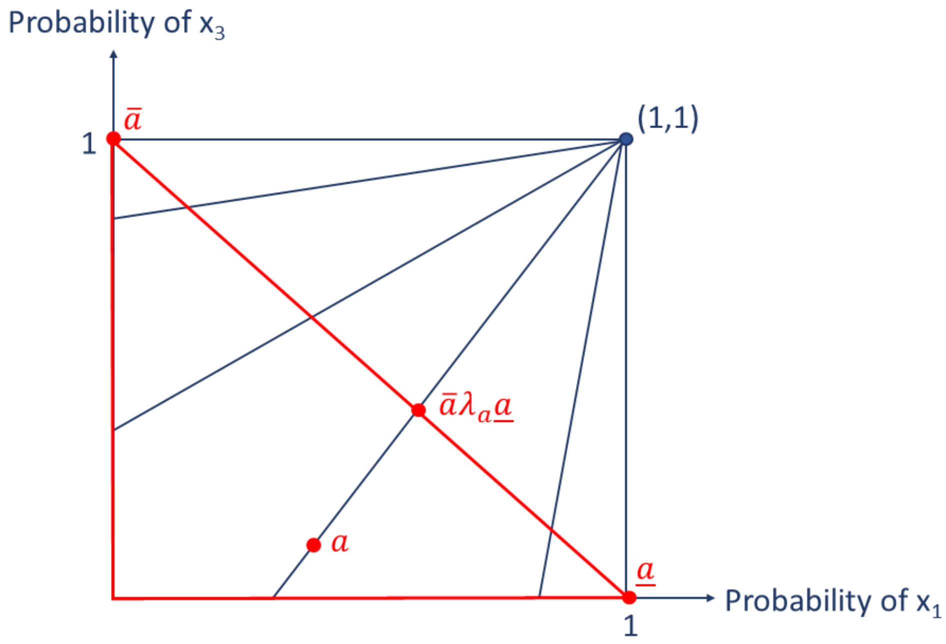

Figure 1: MM triangle for Example 3.1

We wish to decompose CCI into separate conditions that induce mixture-linear utility and Fechnerian structure respectively. In fact, we will do more than this: we will show that CCI can be replaced by two conditions that are *jointly weaker*, but which serve these respective purposes.

Consider the following two axioms, each of which is implied by CCI:

Axiom 5 (Weak Independence) *For any $a, b, c \in A$,*

$$P\left(a\frac{1}{2}c, b\frac{1}{2}c\right) > \frac{1}{2} \quad \Rightarrow \quad \min\left\{P\left(a, a\frac{1}{2}b\right), \; P\left(a\frac{1}{2}b, b\right)\right\} > \frac{1}{2}.$$

Axiom 6 (Stochastic Symmetry) *For any $a, b \in A$ and any $\lambda \in [0,1]$,*

$$P\left(a, a\lambda b\right) \;=\; P\left(b\lambda a, b\right).$$

Stochastic Symmetry is implied by CCI as follows:

$$P\left(a, a\lambda b\right) = P\left(a\left(1-\lambda\right)a, b\left(1-\lambda\right)a\right) = P\left(a\left(1-\lambda\right)b, b\left(1-\lambda\right)b\right) = P\left(b\lambda a, b\right)$$

where CCI is used for the middle equality. To see that Weak Independence also follows from CCI note that we can use the latter to establish

$$P\left(a, a\frac{1}{2}b\right) = P\left(a\frac{1}{2}a, b\frac{1}{2}a\right) = P\left(a\frac{1}{2}c, b\frac{1}{2}c\right) = P\left(a\frac{1}{2}b, b\frac{1}{2}b\right) = P\left(a\frac{1}{2}b, b\right)$$

(where the second and third equalities use CCI) and hence

$$P\left(a\frac{1}{2}c, b\frac{1}{2}c\right) = P\left(a, a\frac{1}{2}b\right) = P\left(a\frac{1}{2}b, b\right).$$

Weak Independence imposes the following restriction on the base relation for P:

$$a\frac{1}{2}c \succ^P b\frac{1}{2}c \quad \Rightarrow \quad \left[a \succ^P a\frac{1}{2}b \text{ and } a\frac{1}{2}b \succ^P b\right]$$

for any $a, b, c \in A$. This has the flavour of a "betweenness" property (Chew, 1983; Dekel, 1986). However, it can be shown that Weak Independence, together with Balance and WST (which is implied by SST), suffice for $\succsim^P$ to satisfy a standard mixture-independence property.

Lemma 3.2 *Let P be a BCP that satisfies Balance, WST and Weak Independence. Then*

$$a \sim^P b \quad \Rightarrow \quad a\frac{1}{2}c \sim^P b\frac{1}{2}c$$

for any $a, b, c \in A$.

The Weak Independence axiom is also closely related to (and its name motivated by) the Independence axiom of Dagsvik (2008), which requires

$$a\lambda c \succ^P b\lambda c \quad \Rightarrow \quad a \succ^P b$$

for any $a, b, c \in A$ and any $\lambda \in (0, 1)$. Given WST and Continuity, Independence implies Weak Independence.[6].

Axiom 6 is a stochastic analogue of the *symmetry* axiom from SSB utility theory (Fishburn, 1984), hence the name. It implies, in particular, that $a\frac{1}{2}b$ is a *stochastic midpoint* between a and b (Davidson and Marschak, 1959), which is critical to establishing a Fechnerian noise structure.

The following example verifies that the conjunction of Axioms 5 and 6 is *strictly* weaker than CCI.

[6]Independence implies that

$$a \sim^P b \quad \Rightarrow \quad a\lambda c \sim^P b\lambda c$$

for any $a, b, c \in A$ and any $\lambda \in (0, 1)$. Given WST (which is implied by SST) and Continuity, it follows that Independence implies Weak Independence – see Fishburn's (1982) proof of H4 from B1-B3 (*ibid.*, pp.16-17).

Example 3.2 *Let A be the unit simplex in $\mathbb{R}^3$ (endowed with the usual mixture operation), which we interpret as the set of all lotteries over some fixed set $X = \{x_1, x_2, x_3\}$ of "prizes". Let $\geq^*$ be the following lexicographic binary relation on the simplex:*

$$a \geq^* b \quad \Leftrightarrow \quad [a_3 > b_3 \text{ or } (a_3 = b_3 \text{ and } a_1 \leq b_1)] \, .$$

(Once again, imagine that x_3 is the best prize and x_1 the worst.) Note that $\geq^*$ is a linear order (i.e., complete, antisymmetric and transitive). Let $>^*$ denote the asymmetric part of $\geq^*$, so

$$a >^* b \quad \Leftrightarrow \quad [a_3 > b_3 \text{ or } (a_3 = b_3 \text{ and } a_1 < b_1)] \, .$$

For any $a, b \in A$ with $a \neq b$, we have $a >^* b$ or $b >^* a$ (but not both).[7] *Furthermore:*

$$a >^* b \quad \Leftrightarrow \quad a\lambda c >^* b\lambda c \tag{3}$$

for any $a, b, c \in A$ and any $\lambda \in (0, 1]$. Finally, given any $a, b \in A$ with $a \neq b$, let $D(a, b)$ denote the Euclidean length of the longest line segment that passes through a and b and remains entirely within the simplex. (Think of $D(a, b)$ as the "width" of the simplex along the line through a and b.) Now define P as follows:

$$P(a, b) \;=\; \begin{cases} \frac{1}{2} & \text{if } a = b \\[2ex] \frac{1}{2} + \frac{1}{2}\left(\frac{\|a-b\|}{D(a,b)}\right) & \text{if } a >^* b \\[2ex] \frac{1}{2} - \frac{1}{2}\left(\frac{\|a-b\|}{D(a,b)}\right) & \text{if } b >^* a \end{cases}$$

It is easy to check that P is a BCP and that

$$P(a, b) > \frac{1}{2} \quad \Leftrightarrow \quad a >^* b \tag{4}$$

To see that P satisfies Axiom 5, use (4) and the fact that (3) implies

$$a\lambda c >^* b\lambda c \quad \Leftrightarrow \quad a >^* b \quad \Leftrightarrow \quad a >^* a\mu b \quad \Leftrightarrow \quad a\eta b >^* b \tag{5}$$

for any $a, b, c \in A$ and any $\lambda, \mu, \eta \in [0, 1]$. That P also satisfies Axiom 6 follows from (5) and two further observations: first, that

$$\|a - a\lambda b\| = (1 - \lambda)\|a - b\| = \|b\lambda a - b\| \, ;$$

[7] Recall that $a \neq b$ implies that $a \in A \subseteq \mathbb{R}^3$ and $b \in A \subseteq \mathbb{R}^3$ differ in at least two components.

and second, that all the vectors in $\{a, b, a\lambda b, b\lambda a\}$ are collinear when $a \neq b$, so $D(a, a\lambda b) = D(b\lambda a, b)$. However, P violates CCI: if $a = \left(\frac{1}{2}, 0, \frac{1}{2}\right)$, $b = c = (0,1,0)$ and $d = (1,0,0)$, then $a >^* b$ and for any $\lambda \in (0,1)$ we have

$$P(a\lambda c, b\lambda c) < 1 = P(a\lambda d, b\lambda d).$$

The MM triangle in Figure 2 illustrates: the probability of x_1 is measured on the horizontal axis and the probability of x_3 on the vertical.

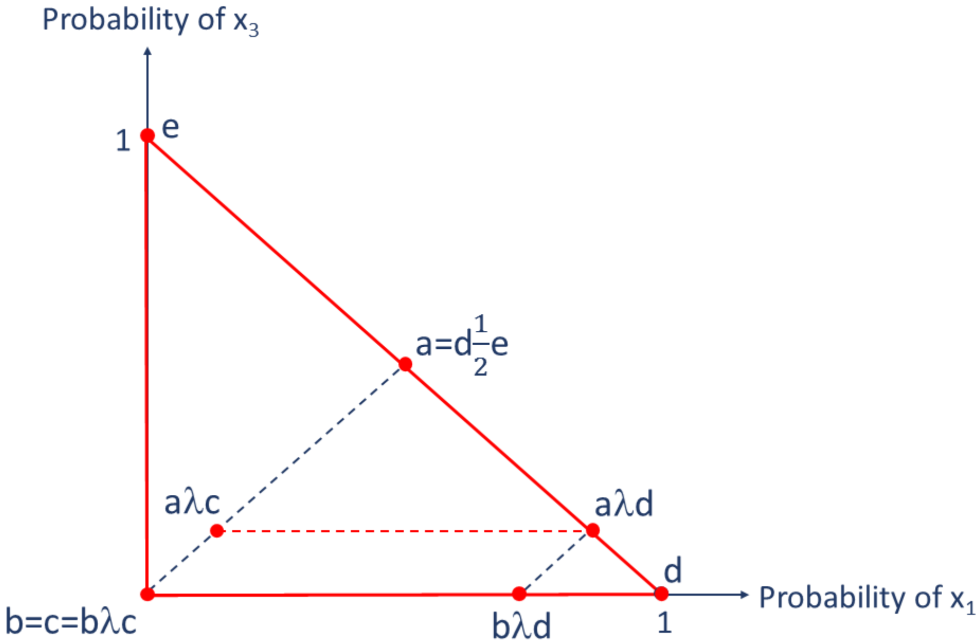

Figure 2: MM triangle for Example 3.2. Note that $a >^* b$.

The next two theorems identify the implications of successively adding Axiom 5 and then Axiom 6 to Axioms 1-3.

Theorem 3.2 *Let P be a binary choice probability. Then P is strictly scalable by some (u, F) with u mixture-linear iff P satisfies Axioms 1-3 and 5.*

Theorem 3.3 *The binary choice probability P has a strict Fechner model (u, G) with u mixture-linear iff P satisfies Axioms 1-3 and 5-6.*

The following corollary is immediate:

Corollary 3.1 *Let P be a binary choice probability such that P is strictly scalable by some (u, F) with u mixture-linear. Then P has a strict Fechner model (v, G) with v mixture-linear iff P satisfies Stochastic Symmetry.*

The proofs of Theorems 3.2 and 3.3 construct the desired representations in a modular fashion. Ryan (2018b) shows that necessary and sufficient conditions for strict scalability are Axioms 1-2 together with the existence of a weak utility for P. As noted above, Axioms 1-3 together with Weak Independence ensure that P has a mixture-linear weak utility. This establishes the "if" part of Theorem 3.2. We then show that Stochastic Symmetry suffices to ensure that F depends only on utility differences, giving the Fechnerian representation in Theorem 3.3.

3.2 Models with strict monotonicity

Using a similar modular structure, we can also axiomatise the important classes of *simply scalable* BCPs (Tversky and Russo, 1969) and BCPs with a *strong utility* (Debreu, 1958; Marschak, 1960). These are subsets of the strictly scalable BCPs and the BCPs with strict Fechner models, respectively. They require *strict* monotonicity of the function (F or G) that converts weak utilities into choice probabilities. These are classical models in the literature – binary logit (Luce, 1959) is the best known example of a strong utility model – though they exclude the possibility of choice being *certain* (i.e., $P(a, b) \in \{0, 1\}$) unless choosing between the most "extreme" alternatives.

Definition 3 *A binary choice probability P is* ***simply scalable*** *iff there exists a function $u : A \to \mathbb{R}$, and a function $F : u(A) \times u(A) \to [0, 1]$ that is strictly increasing (respectively, strictly decreasing) in its first (respectively, second) argument and satisfies $F(x, y) + F(y, x) = 1$ for all $x, y \in u(A)$, such that*

$$P(a, b) = F(u(a), u(b))$$

for all $a, b \in A$. In this case, we say that P is simply scalable by (u, F).

Definition 4 *A binary choice probability P has a* ***strong Fechner model*** *iff there exists a function $u : A \to \mathbb{R}$ and a strictly increasing function $G : \Gamma \to [0, 1]$, where $\Gamma = u(A) - u(A)$, such that $G(x) + G(-x) = 1$ for all $x \in \Gamma$ and*

$$P(a, b) = G(u(a) - u(b))$$

for all $a, b \in A$. In this case, we say that (u, G) is a ***strong Fechner model*** *for P.*

If (u,F) is a strong Fechner model for P then

$$P(a,b) \geq P(c,d) \quad \Leftrightarrow \quad u(a) - u(b) \geq u(c) - u(d) \tag{6}$$

for any $a,b,c,d \in A$. A function $u : A \to \mathbb{R}$ that satisfies (6) for all $a,b,c,d \in A$ is called a *strong utility* for P. Note that P has a strong Fechner model iff it has a strong utility. It is also obvious that if P has a strong utility then it is simply scalable. Moreover, if P is simply scalable by (u,F) then

$$P(a,b) \geq \frac{1}{2} \quad \Leftrightarrow \quad F(u(a), u(b)) \geq F(u(b), u(b)) \quad \Leftrightarrow \quad u(a) \geq u(b)$$

for any $a,b \in A$, so u is a weak utility for P and P is therefore strictly scalable by (u,F). Likewise, if (u,G) is a strong Fechner model for P then it is also a strict Fechner model for P.

To characterise these models we require a strengthening the SST condition:

Axiom 7 (SSST) *For any* $a,b,c \in A$,

$$\min\{P(a,b),\ P(b,c)\} \ \geq [>] \ \frac{1}{2} \quad \Rightarrow \quad P(a,c) \ \geq [>] \ \max\{P(a,b),\ P(b,c)\}.$$

This axiom plays a decisive role in Russo and Tversky (1969) and goes by various names. Tversky and Russo themselves refer it as "strong stochastic transitivity" but this term is now firmly affixed in the literature to the weaker concept defined by Axiom 2. Fishburn (1973) calls it "strict stochastic transitivity", abbreviated "SSST", but this invites confusion in the present context: SSST implies SST while every strong Fechner model is a strict Fechner model. Roberts (1971) calls it the "strong version of strong stochastic transitivity", with acronym SSST. This name would suit our purpose but is a bit cumbersome, so we will refer to Axiom 7 simply by the SSST acronym (consistently with both Fishburn and Roberts).

Theorem 3.4 *Let P be a binary choice probability. Then P is simply scalable by some (u,F) with u mixture-linear iff P satisfies Axioms 1, 3, 5 and 7.*

Theorem 3.5 *The binary choice probability P has a mixture-linear strong utility iff P satisfies Axioms 1, 3 and 5-7.*

From these we immediately deduce:

Corollary 3.2 *Let P be a binary choice probability such that P is simply scalable by some (u,F) with u mixture-linear. Then P has a mixture-linear strong utility iff P satisfies Stochastic Symmetry.*

Tversky and Russo (1969) showed that Axioms 1 and 7 are necessary and sufficient for P to be simply scalable. In Ryan (2018b) it is shown that P is strictly scalable iff it satisfies Axioms 1-2 and there exists a weak utility for P. It is therefore intuitive that the behavioural difference between the model in Theorem 3.2 and that in Theorem 3.4 should be characterised by the difference between SST and SSST.

3.3 Continuous models

Note that P may possess a mixture-linear strong utility yet not possess any strong Fechner model (u, G) in which u is mixture linear and G is *continuous*.

Example 3.3 *Suppose* $A = [0, 1]$ *and*

$$P(a, b) = \begin{cases} \frac{1}{4} - \frac{1}{4}(b - a) & \text{if } a < b \\ \frac{1}{2} & \text{if } a = b \\ \frac{3}{4} + \frac{1}{4}(a - b) & \text{if } a > b \end{cases}$$

Hence, the range of P *is* $\left[0, \frac{1}{4}\right) \cup \left\{\frac{1}{2}\right\} \cup \left(\frac{3}{4}, 1\right]$. *If* $u : A \to \mathbb{R}$ *is mixture-linear, then* $u(A)$ *is an interval, and so is* $\Gamma = u(A) - u(A)$, *which means that* $G(\Gamma)$ *must also be an interval for any continuous and stictly increasing* G. *It follows that* P *cannot have a strong Fechner model* (u, G) *in which* u *is mixture linear and* G *is continuous. However, the identity function is a mixture-linear strong utility for* P, *as the reader may easily verify.*

Strong Fechner models with G continuous are obviously of practical significance. Dagsvik (2008) determines necessary and sufficient conditions for P to possess a strong Fechner model (u, G) with u mixture-linear and G continuous. We obtain a quite different axiomatisation of this class (Theorem 3.9 below). To ensure the continuity of G, we need one additional restriction on P: Debreu's (1958) *solvability* condition.

Axiom 8 (Solvability) *For any* $a, b, c \in A$ *and any* $\rho \in [0, 1]$ *if*

$$P(a, b) \geq \rho \geq P(a, c)$$

then $P(a, d) = \rho$ *for some* $d \in A$.

In fact, if P is strictly (respectively, simply) scalable with a mixture-linear utility scale, then solvability is also necessary and sufficient to ensure that P is strictly (respectively, simply) scalable by some (u, F) with u mixture-linear and F continuous in each argument.[8]

[8] It follows easily that F is therefore (jointly) continuous, as observed by Debreu (1958),

Theorem 3.6 *Let P be a binary choice probability. There exists a mixture-linear u such that P is strictly scalable by (u, F) for some F that is continuous in each argument iff P satisfies Axioms 1-3, 5 and 8.*

Theorem 3.7 *The binary choice probability P has a strict Fechner model (u, G) in which u is mixture linear and G continuous iff P satisfies Axioms 1-3, 5-6 and 8.*

Theorem 3.8 *Let P be a binary choice probability. There exists a mixture-linear u such that P is simply scalable by (u, F) for some F that is continuous in each argument iff P satisfies Axioms 1, 3, 5 and 7-8.*

Theorem 3.9 *The binary choice probability P has a strong Fechner model (u, G) in which u is mixture linear and G continuous iff P satisfies Axioms 1, 3 and 5-8.*

Theorems 3.3 and 3.9 characterise the models in Blavatskyy (2008) and Dagsvik (2008) respectively. The common axiomatic core consists of Axioms 1, 3, 5 and 6. For Blavatskyy's model we add SST (Axiom 2), while for Dagsvik's we add SSST and Solvability (Axioms 7 and 8). Strengthing SST to SSST ensures that the function converting utility differences to choice probabilities is *strictly* increasing, while solvability ensures it is continuous.

While this axiomatic structure allows us to construct the various representations in conveniently modular pieces, by the time we reach Dagsvik's model there are six axioms in total. It is natural to wonder whether the two varieties of continuity axiom – Axioms 3 and 8 – might be consolidated into a single axiom that ensures continuity of u and F simultaneously? With a slight strengthening of Weak Independence, such economy is possible.

Axiom 5′ *For any $a, b, c \in A$ and any $\lambda \in (0, 1]$,*

$$P(a\lambda c, b\lambda c) > \frac{1}{2} \quad \Rightarrow \quad \min\left\{P\left(a, a\frac{1}{2}b\right), \ P\left(a\frac{1}{2}b, b\right)\right\} > \frac{1}{2}.$$

It is straightforward to see Axiom 5′ is also an implication of CCI. Moreover, each of our results remains valid if Axiom 5 is replaced with Axiom 5′, since the latter is clearly necessary for the existence of a mixture-linear weak utility.

Given Axiom 5′ we may replace Axioms 3 and 8 in Theorem 3.9 with the following strengthened form of Axiom 8:

Axiom 8′ (Mixture Solvability) *For any $a, b, c \in A$ and any $\rho \in [0, 1]$ if*

$$P(a, b) \geq \rho \geq P(a, c)$$

then $P(a, b\lambda c) = \rho$ for some $\lambda \in [0, 1]$.

	weak	**strict**
scalable	*SST* *Weak Independence* (Theorem 3.2)	*SSST* *Weak Independence* (Theorem 3.4)
Fechnerian	*SST* *Weak Independence* *Stochastic Symmetry* (Theorem 3.3)	*SSST* *Weak Independence* *Stochastic Symmetry* (Theorem 3.5)

Table 1: Models without continuity of F or G

This mixture solvability condition is taken from Ryan (2018a). We may now state our final representation result:

Theorem 3.10 *The binary choice probability P has a strong Fechner model (u, G) in which u is mixture linear and G continuous iff P satisfies Axioms 1, $5'$, 6, 7 and $8'$.*

4 Concluding remarks

Tables 1 and 2 summarise the main results of the paper, and highlight the modular nature of our axiomatisations. Table entries give the axiomatisation of all eight models discussed in the paper. Each model has a mixture-linear weak utility function; those in the first row of each table are the scalable models, those in the second Fechnerian; columns indicate the monotonicity of F or G as appropriate: weak versus strict. The tables are distinguished by whether or not the models impose a continuous transformation of weak utilities into choice probabilities: continuity is imposed in Table 2 but not in Table 1. Since Axioms 1 and 3 are standard and assumed for every model, they are excluded from the table; we only list the *additional* axioms required to characterise the respective models.

Tables 1 and 2 emphasise the modularity of our axiomatisations. To move from left to right along a row, we replace SST with SSST; moving down a column, we add Stochastic Symmetry; and to move from Table 1 to the corresponding entry in Table 2, we add Solvability.

It is also noteworthy that the prior literature only offered axiomatisations of two of these eight models: the lower left-hand model in Table 1 (Blavatskyy, 2008) and the lower right-hand model in Table 2 (Dagsvik, 2008). Not only do our results characterise the other six models; they also provide new axiomatisations of the existing two.[9]

[9]Dagsvik (2015) also provides alternative characterisations of these two models, though of a very different nature to the axiomatisations given here.

	weak, continuous	**strict, continuous**
scalable	*SST* *Weak Independence* *Solvability* (Theorem 3.6)	*SSST* *Weak Independence* *Solvability* (Theorem 3.8)
Fechnerian	*SST* *Weak Independence* *Stochastic Symmetry* *Solvability* (Theorem 3.7)	*SSST* *Weak Independence* *Stochastic Symmetry* *Solvability* (Theorem 3.9)

Table 2: Models with continuity of F or G

In our proofs, Weak Independence is used to ensure the existence of a mixture-linear weak utility, then Stochastic Symmetry to underwrite the Fechnerian structure of noise. This division of labour allows us to characterise both strict and strong Fechner models, as well as strict and simple scalability, all within the same basic proof structure. As noted in the Introduction, our decomposition of CCI into Weak Independence and Stochastic Symmetry may have additional value for the experimental testing of models that embed EU in a Fechnerian noise structure. If CCI fails but Stochastic Symmetry is supported by the data, then von Neumann and Morgenstern are rejected but not Fechner.

Loomes and Sugden (1998, pp.594-5) themselves blame von Neuman and Morgenstern for the failure of CCI in their experimental data, but based on quite different reasoning. They detect what they call a *bottom edge effect*, which is inconsistent with any known stochastic form of EU. This conclusion is based on casual (albeit convincing) empiricism rather than formal testing. Our approach offers a formal avenue for confirming their conjecture.[10]

This testing strategy has the further advantage that both CCI and Stochastic Symmetry can be tested on *aggregate* data; that is, on data which pools the choices of multiple subjects. If multiple subjects make choices from the binary choice sets $\{a\lambda c, b\lambda c\}$ and $\{a\lambda d, b\lambda d\}$, then CCI implies that the proportion of times that $a\lambda c$ is chosen over $b\lambda c$ (across subjects) should match the proportion of times that $a\lambda d$ is chosen over $b\lambda d$, even if the common choice probability $P(a\lambda c, b\lambda c) = P(a\lambda d, b\lambda d)$ differs from subject to subject. As is well known (Ballinger and Wilcox, 1997) many other properties of stochastic choice models are not similarly robust to aggregation over heterogeneous individuals; properties from individual choice probabilities are not inherited by aggregate choice proportions.

It is also worth noting that nothing in Corollaries 3.1 and 3.2 restricts their domain to risk. Since A need only be a mixture set, it might, for example, be a set of Anscombe-

[10] Note, however, that violation of Axiom 6 would not invalidate their hypothesis.

Aumann acts in a domain of uncertainty. On this domain, *subjective expected utility (SEU)* is mixture-linear. It follows that Stochastic Symmetry is also the bridge from scalability with respect to an SEU "scale" to SEU maximisation with Fechnerian noise, at least within an Anscombe-Aumann environment.

Finally, let us observe that our scalability results (Theorems 3.2 and 3.4) could easily be extended to multinomial choice. The only substantive change would be the replacement of SSST (Axiom 7) with Tversky's (1972) *order independence* axiom, and SST (Axiom 2) with the *multinomial weak substitutability* condition of Ryan (2018b). Axiom 1 would also need to be replaced with the obvious multinomial generalisation, while Axiom 3 and Weak Independence require no modification other than translation into suitable notation for multinomial choice. See Ryan (2018b) for further details.

References

Ballinger, T. P. and N.T. Wilcox (1997) "Decisions, Error and Heterogeneity" *The Economic Journal* **107(443)**, 1090-1105.

Blavatskyy, P.R. (2008) "Stochastic Utility Theorem" *Journal of Mathematical Economics* **44**, 1049–1056.

Blavatskyy, P. R. and G. Pogrebna (2010) "Models of Stochastic Choice and Decision Theories: Why both are Important for Analyzing Decisions" *Journal of Applied Econometrics* **25(6)**, 963-986.

Buschena, D., and D. Zilberman (2000) "Generalized Expected Utility, Heteroscedastic Error, and Path Dependence in Risky Choice" *Journal of Risk and Uncertainty* **20(1)**, 67-88.

Chew, S.H. (1983) "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox" *Econometrica* **51(4)**, 1065-1092.

Conte, A., J.D. Hey and P.G. Moffatt (2011) "Mixture Models of Choice under Risk" *Journal of Econometrics* **162(1)**, 79-88.

Dagsvik, J.K. (2008) "Axiomatization of Stochastic Models for Choice Under Uncertainty" *Mathematical Social Sciences* **55**, 341–370.

Dagsvik, J. K. (2015) "Stochastic Models for Risky Choices: A Comparison of Different Axiomatizations" *Journal of Mathematical Economics* **60**, 81-88.

Davidson, D. and J. Marschak (1959) "Experimental Tests of a Stochastic Decision Theory" in C.W. Churchman and P. Ratoosh (eds), *Measurement: Definitions and Theories*. John Wiley and Sons: New York.

Debreu, G. (1958) "Stochastic Choice and Cardinal Utility" *Econometrica* **26(3)**, 440–444.

Dekel, E. (1986) "An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom" *Journal of Economic Theory* **40(2)**, 304-318.

Fishburn, P. C. (1973) "Binary Choice Probabilities: On the Varieties of Stochastic Transitivity" *Journal of Mathematical Psychology* **10(4)**, 327-352.

Fishburn, P.C. (1982) *The Foundations of Expected Utility*, D. Reidel Publishing: Dordrecht.

Fishburn, P. C. (1984) "SSB Utility Theory: An Economic Perspective" *Mathematical Social Sciences* **8(1)**, 63-94.

Hey, J. D. (1995) "Experimental Investigations of Errors in Decision Making under Risk" *European Economic Review* **39(3-4)**, 633-640.

Hey, J.D. (2014) "Choice under Uncertainty: Empirical Methods and Experimental Results" in M. J. Machina, W. K. Viscusi (eds), *Handbook of the Economics of Risk and Uncertainty*, Volume 1. Oxford: North Holland.

Loomes, G. and R. Sugden (1998) "Testing Different Stochastic Specifications of Risky Choice" *Economica* **65(260)**, 581-598.

Luce, R.D. (1959) *Individual Choice Behavior: A Theoretical Analysis*, John Wiley and Sons: New York.

Marschak, J. (1960) "Binary Choice Constraints and Random Utility Indicators" in K.J. Arrow, S. Karlin and P. Suppes (eds) *Mathematical Methods in the Social Sciences*. Stanford University Press, Stanford, CA.

Regenwetter, M., J. Dana and C.P. Davis-Stober (2011) "Transitivity of Preferences" *Psychological Review* **118(1)**, 42-56.

Roberts, F. S. (1971) "Homogeneous Families of Semiorders and the Theory of Probabilistic Consistency" *Journal of Mathematical Psychology* **8(2)**, 248-263.

Ryan, M. J. (2015) "A Strict Stochastic Utility Theorem" *Economics Bulletin* **35(4)**, 2664-2672.

Ryan, M. J. (2018a) “Uncertainty and Binary Stochastic Choice” *Economic Theory* **65(3)**, 629-662.

Ryan, M. J. (2018b) “Strict Scalability of Choice Probabilities” *Journal of Mathematical Psychology* **18**, 89-99.

Tversky, A. (1972), “Choice by Elimination,” *Journal of Mathematical Psychology* **9**, 341–367.

Tversky, A. and J.E. Russo (1969) “Substitutability and Similarity in Binary Choices” *Journal of Mathematical Psychology* **6**, 1–12.

Appendix

Proof of Lemma 3.1: Since u and v are both weak utilities for P, there is a strictly increasing function $h : u(A) \to \mathbb{R}$ such that $v = h \circ u$. Thus

$$F(x, y) = G(h(x) - h(y)) \tag{7}$$

for all $x, y \in u(A)$. It remains to show that h is continuous.

Since $u(A)$ is a non-degenerate interval and h is strictly increasing, v is non-constant. As noted in Ryan (2015), if (v, G) is a strict Fechner model for P with v non-constant, then G cannot be constant on any open interval containing 0. We will show that this implies the continuity of h.

Suppose, to the contrary, that h is *not* continuous. Then there is some $\{x^n\}_{n=1}^{\infty} \subseteq u(A)$ with $x^n \to \hat{x} \in u(A)$ as $n \to \infty$ but $h(x^n)$ does not converge to $h(\hat{x})$. That is, there exists some $\varepsilon > 0$ such that either $h(x^n) - h(\hat{x}) \geq \varepsilon$ infinitely often, or $h(x^n) - h(\hat{x}) \leq -\varepsilon$ infinitely often. We only consider the former case, as the latter may be handled similarly. Since $x^n \to \hat{x} \in u(A)$ and F is continuous in each argument, we have

$$F(x^n, \hat{x}) \to F(\hat{x}, \hat{x}) = \frac{1}{2} \tag{8}$$

as $n \to \infty$ (where we have used the fact that $F(x, y) + F(y, x) = 1$). Since $h(x^n) - h(\hat{x}) \geq \varepsilon$ infinitely often and G is non-decreasing, it follows from (7) that:

$$F(x^n, \hat{x}) = G(h(x^n) - h(\hat{x})) \geq G(\varepsilon) \text{ for infinitely many } n \tag{9}$$

Moreover, since G is non-decreasing and satisfies $G(x) + G(-x) = 1$, we have:

$$G(\varepsilon) \geq G(0) = \frac{1}{2}.$$

From (8) and (9) we therefore conclude that $G(\varepsilon) = G(0)$; hence

$$G(z) = \frac{1}{2}$$

for all $z \in [0, \varepsilon]$ and

$$G(z) = 1 - G(-z) = \frac{1}{2}$$

for all $z \in [0, -\varepsilon]$. Thus G is constant on $(-\varepsilon, \varepsilon)$, which is the desired contradiction. $\square$

Proof of Lemma 3.1: Suppose, to the contrary, that $a \sim^P b$ but

$$a\frac{1}{2}c \succ^P b\frac{1}{2}c.$$

Then

$$P\left(a\frac{1}{2}c, b\frac{1}{2}c\right) > \frac{1}{2}.$$

Using Axiom 5 we have

$$\min\left\{P\left(a, a\frac{1}{2}b\right), P\left(a\frac{1}{2}b, b\right)\right\} > \frac{1}{2} \tag{10}$$

so $P(a,b) \geq \frac{1}{2}$ by WST. If $P(a,b) = \frac{1}{2}$ then Balance gives $P(b,a) = \frac{1}{2}$ and we obtain

$$P\left(b, a\frac{1}{2}b\right) \geq \frac{1}{2}$$

from WST and (10). Using Balance again we have

$$P\left(a\frac{1}{2}b, b\right) \leq \frac{1}{2}$$

which contradicts (10). We can therefore rule out $P(a,b) = \frac{1}{2}$ so we must have $P(a,b) > \frac{1}{2}$. But this contradicts $a \sim^P b$. $\square$

The following will be useful in the sequel:

Lemma 4.1 *Let (A, P) be simply (respectively, strictly) scalable through (u, F). If $h : u(A) \to \mathbb{R}$ is strictly increasing and $\hat{u} = h \circ u$, then there exists an $\hat{F}$ such that (A, P) is simply (respectively, strictly) scalable through $\left(\hat{u}, \hat{F}\right)$.*

Proof: The condition

$$\hat{F}(x,y) \;=\; F\left(h^{-1}(x), h^{-1}(y)\right)$$

determines a well-defined function $\hat{F} : \hat{u}(A) \times \hat{u}(A) \to [0,1]$ which shares the same monotonicity properties as F. Moreover, if u represents $\succsim^P$ then so does $\hat{u}$. $\square$

Proof of Theorem 3.2. Suppose P satisfies Axioms 1-3 and 5.

We first show that $\succsim^P$ has a mixture-linear representation. The argument closely follows Step 1 in the proof of Corollary 2.1 in Ryan (2015). The base relation is complete by Axiom 1 and transitive by strong stochastic transitivity (Axiom 2). Using Axiom 3 we deduce that the sets

$$\left\{\lambda \in [0,1] \;\middle|\; a\lambda b \succsim^P c\right\}$$

and

$$\left\{\lambda \in [0,1] \;\middle|\; c \succsim^P a\lambda b\right\}$$

are closed for any $a, b, c \in A$. By Theorem 1 in Fishburn (1982, Chapter 2), it suffices to show that $\succsim^P$ satisfies the following independence condition: for any $a, b, c \in A$

$$a \sim^P b \quad \Rightarrow \quad a\frac{1}{2}c \sim^P b\frac{1}{2}c.$$

But this follows from Lemma 3.2, using the fact that SST implies WST.

Lemma 11 and Theorem 14 in Ryan (2018b) now imply that P is strictly scalable. Using Lemma 4.1 and the fact that $\succsim^P$ has a mixture-linear utility representation, we deduce that P is strictly scalable by some (u, F) with u mixture-linear.

Conversely, suppose P is strictly scalable by (u, F) with u mixture-linear. Axiom 1 follows from the fact that $F(x,y) + F(y,x) = 1$. Axiom 2 (SST) follows from the facts that u represents $\succsim^P$ and the monotonicity properties of F: if $P(a,b) \geq \frac{1}{2}$ and $P(b,c) \geq \frac{1}{2}$ then $u(a) \geq u(b) \geq u(c)$ so

$$F(u(a), u(c)) \geq \max\left\{F(u(a), u(b)),\; F(u(b), u(c))\right\}.$$

To verify Continuity (Axiom 3) we use the mixture-linearity of u and the fact that u represents $\succsim^P$ to deduce

$$P(a\lambda b, c) \geq \frac{1}{2} \quad \Leftrightarrow \quad u(a\lambda b) \geq u(c) \quad \Leftrightarrow \quad \lambda\left[u(a) - u(b)\right] \geq \left[u(c) - u(b)\right]$$

and

$$P(a\lambda b, c) \leq \frac{1}{2} \quad \Leftrightarrow \quad u(a\lambda b) \leq u(c) \quad \Leftrightarrow \quad \lambda\left[u(a) - u(b)\right] \leq \left[u(c) - u(b)\right].$$

Weak Independence (Axiom 5) also follows from the mixture-linearity of u and the fact that u represents $\succsim^P$: if $a\frac{1}{2}c \succ^P b\frac{1}{2}c$ then

$$\frac{1}{2}u(a)+\frac{1}{2}u(c)=u\left(a\frac{1}{2}c\right)>u\left(b\frac{1}{2}c\right)=\frac{1}{2}u(b)+\frac{1}{2}u(c)$$

and hence $u(a)>u(b)$. Therefore

$$u(a)>\frac{1}{2}u(a)+\frac{1}{2}u(b)>u(b)$$

so $a \succ^P a\frac{1}{2}b \succ^P b$. $\square$

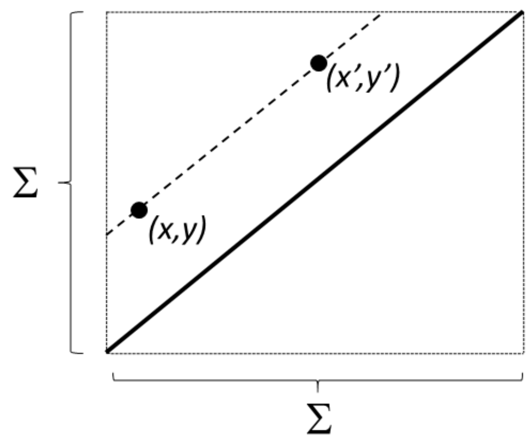

Figure 3: Domain of F

Proof of Theorem 3.3. Suppose P satisfies Axioms 1-3 and 5-6. By the previous theorem, P is strictly scalable by some (u,F) with u mixture-linear. If remains to show that $F(x,y)=F(x',y')$ whenever $x,y,x',y'\in u(A)$ with $x-y=x'-y'$. Let $\Sigma=u(A)$ and recall that Σ is an interval, possibly unbounded. Figure 3 illustrates the domain of F. Since $F(x,y)+F(y,x)=1$ for any $x,y\in\Sigma$, the contours of F are symmetric about the 45 degree line (diagonal) in Figure 3. Moreoever, $F(x,x)=\frac{1}{2}$ so any two points on the 45 degree line are contained within the same contour of F. It therefore suffices to show that any two distinct points on a line parallel to, and above, the 45 degree line occupy the same contour of F. Let (x,y) and (x',y') be two such points, so $y-x=y'-x'>0$.

Without loss of generality (WLOG), we assume that $x' > x$ (as in Figure 3). It follows that $\{y, x'\} \subseteq (x, y')$ with $y = \lambda x + (1-\lambda) y'$ and $x' = \lambda y' + (1-\lambda) x$ for some $\lambda \in (0,1)$.[11] Let $a, b, a', b' \in A$ be such that $x = u(a)$, $y = u(b)$, $x' = u(a')$ and $y' = u(b')$. Then Stochastic Symmetry and the mixture-linearity of u imply

$$\begin{aligned} F(x,y) = F(x, \lambda x + (1-\lambda) y') &= P(a, a\lambda b') \\ &= P(b'\lambda a, b') = F(\lambda y' + (1-\lambda) x, y') = F(x', y') \end{aligned}$$

as required.

Conversely, suppose P has a mixture-linear strong utility, u. Since P is therefore strictly scalable, it suffices, given what was established in Theorem 3.2, to verify Axiom 6 (Stochastic Symmetry). This follows straightforwardly from the mixture-linearity of u:

$$\begin{aligned} P(a, a\lambda b) = P(b\lambda a, b) \quad &\Leftrightarrow \quad u(a) - u(a\lambda b) = u(b\lambda a) - u(b) \\ &\Leftrightarrow \quad (1-\lambda)[u(a) - u(b)] = (1-\lambda)[u(a) - u(b)] \end{aligned}$$

□

Proof of Theorem 3.4. Suppose P satisfies Axioms 1, 3, 5 and 7. The proof of Theorem 3.2 establishes that P has a mixture-linear weak utility. Since P satisfies Balance and SSST, it is well-known that it is simply scalable (Tversky and Russo, 1969). Hence, from Lemma 4.1 it follows that P is simply scalable by some (u, F) with u mixture-linear.

Conversely, suppose P is simply scalable by (u, F) with u mixture-linear. Since P is therefore also strictly scalable by (u, F), Theorem 3.2 ensures that P satisfies Axioms 1, 3 and 5, as well as SST. The monotonicity properties of F ensure that if $u(a) > u(b) > u(c)$, then

$$F(u(a), u(c)) > \max\{F(u(a), u(b)),\ F(u(b), u(c))\}.$$

Since u represents $\succsim^P$, SSST follows (given SST). □

Proof of Theorem 3.5. The result follows by the same argument as for Theorem 3.3, *mutatis mutandis.* □

Proof of Theorem 3.6. Suppose P satisfies Axioms 1-3, 5 and 8. The proof of Theorem 3.2 establishes that P is strictly scalable by some (u, F) with u mixture-linear. Since $u(A)$ is an interval, if there exists an $a \in A$ such that $F(a, \cdot)$ is not continuous, there must be a gap in the range of $F(a, \cdot)$, since $F(a, \cdot)$ is non-increasing. This would imply a violation of Solvability. Similarly, if there exists an $a \in A$ such that $F(\cdot, a)$ is not continuous, then we can use Axiom 1 (Balance) to obtain another violation of Solvability.

[11] Let $y = \mu x + (1-\mu) y'$ and $x' = \lambda x + (1-\lambda) y'$. Then $y - x = y' - x' > 0$ implies $\mu = 1 - \lambda$.

Conversely, suppose P is strictly scalable by (u, F) with u mixture-linear and F continuous in each argument. Theorem 3.2 ensures that P satisfies Axioms 1-3 and 5. It remains to verify Solvability. If

$$F(u(a), u(b)) \geq \rho \geq F(u(a), u(c))$$

then the properties of u and F ensure that $F(u(a), u(b\lambda c)) = \rho$ for some $\lambda \in [0,1]$: defining $h : [0,1] \to [0,1]$ by $h(\lambda) = F(u(a), u(b\lambda c))$, we see that h is continuous with $h(0) \leq \rho \leq h(1)$, so $h(\lambda) = \rho$ for some $\lambda \in [0,1]$ by the Intermediate Value Theorem. □

Proof of Theorem 3.7. Suppose P satisfies Axioms 1-3, 5-6 and 8. The proof of Theorem 3.3 establishes that P has a strict Fechner model (u, G) with u mixture-linear. Since $u(A)$ is an interval, so is $\Gamma = u(A) - u(A)$. If G is not continuous, there must be a gap in the range of G, since G is non-decreasing. This would imply a violation of Solvability.

Conversely, suppose P has a strict Fechner model (u, G) with u mixture-linear and G continuous. Theorem 3.3 ensures that P satisfies Axioms 1-3 and 5-6. We may verify Solvability using an analogous argument to the one in the proof of Theorem 3.6. □

Proof of Theorem 3.8. This follows by the argument used to prove Theorem 3.6, *mutatis mutandis*, with Theorem 3.4 used in place of Theorem 3.2. □

Proof of Theorem 3.9. This follows by the argument used to prove Theorem 3.7, *mutatis mutandis*, with Theorem 3.5 used in place of Theorem 3.3. □

Proof of Theorem 3.10. The necessity of Axiom 5′ follows by a similar argument to that for the necessity of Axiom 5 in the proof of Theorem 3.2, and the necessity of Mixture Solvability by a similar argument to that for the necessity of Solvability in the proof of Theorem 3.8. It remains to verify the sufficiency of the axioms. For this, it suffices to show that the assumed axioms imply Axiom 3 (continuity), since Axiom 5′ implies Axiom 5 and Mixture Solvability implies Solvability.

We first modify the argument in the proof of Theorem 3.2 to show that $\succsim^P$ satisfies the following independence condition: for any $a, b, c \in A$ and any $\lambda \in (0,1)$:

$$a \succsim^P b \quad \Rightarrow \quad a\lambda c \succsim^P b\lambda c \tag{11}$$

Suppose, to the contrary, that $a \succsim^P b$ and

$$b\lambda c \succ^P a\lambda c$$

for some $\lambda \in (0,1)$. Then Axiom 5′ implies

$$\min\left\{P\left(b, a\frac{1}{2}b\right), P\left(a\frac{1}{2}b, a\right)\right\} > \frac{1}{2}.$$

Applying SSST we have $b \succ^P a$, which is the desired contradiction.

Finally, given the independence property (11) and Mixture Solvability (Axiom 8′), the argument on p.653 of Ryan (2018a) shows that P satisfies Axiom 3 (Continuity), which completes the proof. $\square$