School of Economics

Working Paper Series

Stochastic Expected Utility for Binary Choice: New Representations

Matthew Ryan

2018/06

Stochastic Expected Utility for Binary Choice: New Representations

Matthew Ryan*

July 2018

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. In particular, we are able to decompose 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 has significant advantages for testing the Fechnerian models, as we discuss.

*School of Economics, Auckland University of Technology, New Zealand. Email: `mryan@aut.ac.nz`

1 Introduction

Since John Hey issued his provocative challenge in the mid-1990s, declaring that "one can explain experimental anlayses 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), there has emerged a revisionist literature on the descriptive merits of expected utility (EU). This literature views the evidence through the lens of *probabilistic* models 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" together with 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 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 an increasing 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 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 linear and noise Fechnerian. Should we blame Fechner (as Hey conjectures) or von Neumann and Morgenstern, or both?

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 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 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: a "strict" type, in which binary choice probabilities are expressed as a *non-decreasing* function of EU differences, and a "strong" 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 "strict" model, and Dagsvik (2008), who axiomatised a "strong" model in which choice probabilities are also required to vary *continuously* with EU differences. We provide a new axiomatisation for Dagsvik's model as well (see Theorem 3.7).

Unlike those of Blavatskky (2008) and Dagsvik (2008), our axiomatisations have a modular structure that separates the conditions guaranteeing linear utility from those required for Fechnerian noise structure. Along the way we obtain axiomatisations of binary choice probabilities that are ("strictly" or "simply") *scalable* with respect to a 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.

This paper is a companion to Ryan (2018a), and part of the broader literature on what has been called *Luce's challenge* (Regenwetter, Dana and Davis-Stober, 2011): that is, the challenge of identifying methods for converting axiomatic characterisations of particular classes of utility functions into axiomatic characterisations of probabilistic models of choice that embody the noisy maximisation of some utility function within a given class.

Our new axiomatisations have one further ancillary benefit. Despite the close similarity between their two Fechnerian models, the axiomatisations offered by Blavatskyy (2008) and Dagsvik (2008) are very different. Our results set both models upon a common axiomatic foundation. We show that Dagsvik's model can be obtained from Blavatskyy's axiomatisation by strengthening the *strong stochastic transitivity (SST)* axiom to *strict stochastic transitivity* (compare Definitions 2 and 7 below) and adding Debreu's (1958) *solvability* condition (Theorem 3.7). This axiomatic structure also locates the model differences exactly where one's intuition would expect them to be, as is evident in our proofs: 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.[1]

[1]Dagsvik (2015) provides an alternative path to harmonising the axiomatisation of the two models. He sets them upon a common axiomatic core that comprises a subset of the axioms in Dagsvik (2008, Theorem 4). The distinction between the models is in the notion of probabilistic mixture-independence that is added to this core: if CCI is added we obtain Dagsvik's model; if we add Dagsvik's (2008) *Strong Independence* condition instead, we obtain Blavatskyy's model. This appealing irony is offset by an obvious drawback: the axiomatic distinction clashes with our intuition about how the models differ. The differences are not in their respective degrees of "linearity", so it is puzzling to differentiate the models through a probabilistic independence axiom.

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 theorems.

2 Binary choice probabilities

Let A be a *mixture set* of alternatives, 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 $\lambda \in [0,1]$. Note that 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 $\{a,b\}$. Terms of the form $P(a,a)$ for some $a \in A$ will not be given any behavioural interpretation. 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 $a, b \in A$ (and hence $P(a,a) = \frac{1}{2}$ for any $a \in A$), 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 probablity, 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. Note that $\succsim^P$ 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 $\succsim^P$ (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).[2]

3 Models and representations

Our purpose is to axiomatise various classes of models for BCPs. The following two classes of models were defined in Ryan (2018b) and Ryan (2015) respectively:

[2] A notion of "strong" utility will be introduced shortly.

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 $G : \Gamma \to [0,1]$ satisfying $G(x) + G(-x) = 1$ for all $x \in \Gamma$, where $\Gamma = u(A) - u(A)$,*[3] *such that*

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

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 P must satisfy $P(a,b) + P(b,a) = 1$ if it is to be strictly scalable – this follows from the restriction that $F(x,y) + F(y,x) = 1$.

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 that monotone scalability does not require the function 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 focus on BCPs which are monotone scalable with respect to some weak utility for P; that is, on BCPs that are strictly scalable. This restriction is implicit in Blavatskyy (2008), for example – see Ryan (2015).

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]

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

[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 btrong btochastic 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.

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 [as 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 the following example.

Example 3.1 *Let A be the unit simplex in $\mathbb{R}^3$. 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 Machina-Marschak 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. (Once again, 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 $(1,1)$ 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 a to $(1,1)$ 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, *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 00 (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 there exists some strictly increasing function* $v : [0,1] \to \mathbb{R}$ *and some non-decreasing function* $G : \hat{\Gamma} \to [0,1]$, *where* $\hat{\Gamma} = v([0,1]) - v([0,1])$, *such that*

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

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. Thus:*

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

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*

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

but

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

which contradicts (2).

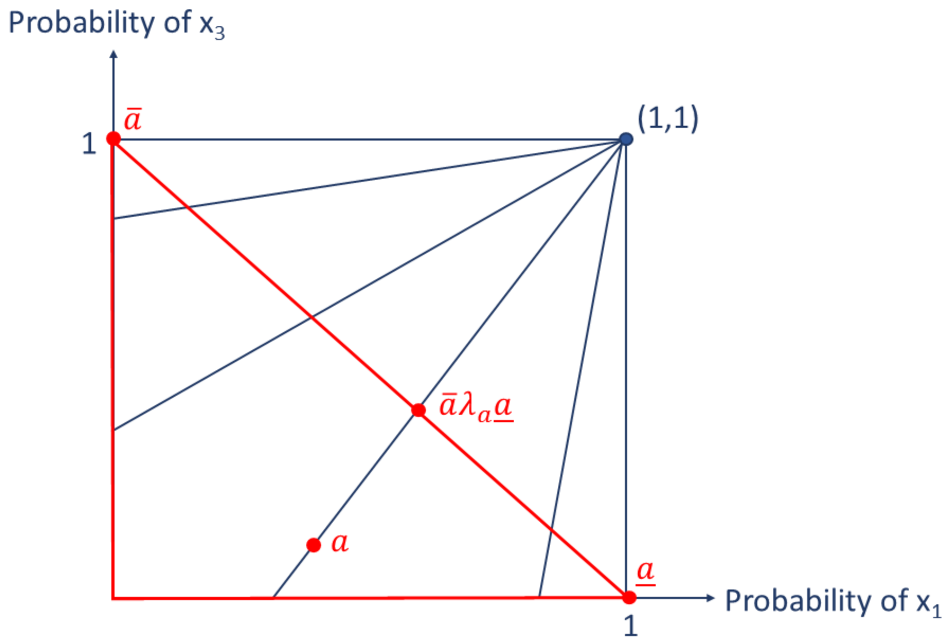

Figure 1: Machina-Marschak triangle for Example 3.1

We wish to decompose CCI into separate conditions that induce linear utility and Fechnerian structure respectively. In fact, we will do more than this: we will show CCI can be replced 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 *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 *For any $a, b \in A$ and any $\lambda \in [0,1]$,*

$$P(a, a\lambda b) = P(b\lambda a, b).$$

To see that CCI implies Axiom 5, note that

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

where the second and third equalities use CCI. Likewise, we obtain Axiom 6 as follows:

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

where CCI is used for the middle equality. Axiom 6 implies, in particular, that $a\frac{1}{2}b$ is a *stochastic mid-point* between a and b (Davidson and Marschak, 1959). It is a probabilistic analogue of the *symmetry* axiom from SSB utility theory (Fishburn, 1984).

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

Example 3.2 *Let A be the unit simplex in $\mathbb{R}^3$, interpreted 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). 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:

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

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

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

Using this latter fact and (3), we see that P satisfies Axiom 5. Since

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

and since all vectors in $\{a, b, a\lambda b, b\lambda a\}$ are collinear when $a \neq b$, P also satisfies Axiom 6. 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 Machina-Marschak 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.

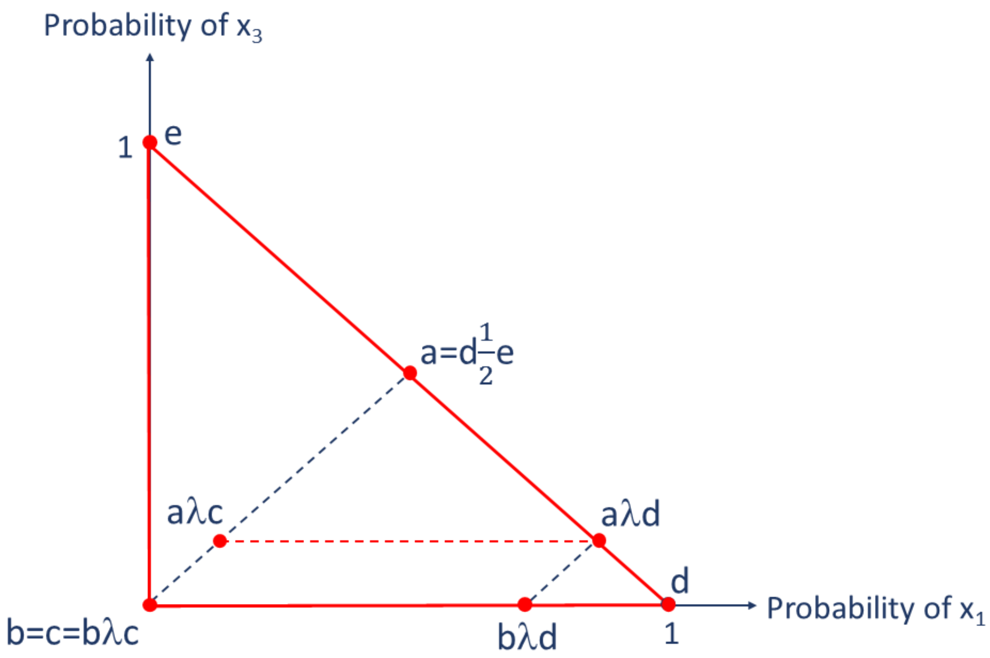

Figure 2: Machina-Marschak triangle for Example 3.2. Note that $a >^* b$.

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 Axiom 6.*

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. Axiom 2 implies the transitivity of $\succsim^P$, so combining this with Axiom 3 (continuity) and Axiom 5, which imposes a mixture-independence condition on $\succsim^P$, we show that there exists a mixture-linear weak utility. This establishes the ("if" part of) Theorem 3.2. We then show that Axiom 6 suffices to ensures that F depends only on utility differences, giving the Fechnerian representation in Theorem 3.3.

Our decomposition of CCI into Axioms 5-6 mirrors Dagsvik's (2015) analogous decomposition of his *strong independence (SI)* axiom into his conditions D4, which imposes a standard mixture-independence restriction on $\succsim^P$, and D5*, which imposes an SI-type condition on the extreme points of the mixture space (the Dirac measures in his simple lottery framework). However, the precise role of D5* is obscured by the fact that Debreu's (1958) *quadruple condition* (Dagsvik's Axiom D1(iii)) also appears amongst Dagsvik's axioms. Indeed, it is the quadruple condition which is used to ensure Fechnerian structure in Dagsvik's proof; condition D5* is used, instead, to establish the mixture-linearity of utility. In our proofs, Axiom 5 is used to ensure the existence of a mixture-linear weak utility and Axiom 6 to underwrite the Fechnerian structure of noise. This gives a very clear division of labour, with transparent behavioural foundations for each aspect of the representation.

Using a similar modular structure, we next 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.

Definition 3 *A binary choice probability P is* ***simply scalable*** *iff there exists a utility 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 utility function $u : A \to \mathbb{R}$ and a strictly increasing function $G : \Gamma \to [0,1]$ satisfying $G(x) + G(-x) = 1$ for all $x \in \Gamma$, where $\Gamma = u(A) - u(A)$, such that*

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

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

$$P(a,b) \geq P(c,d) \quad \Leftrightarrow \quad u(a) - u(b) \geq u(c) - u(d) \tag{4}$$

for any $a, b, c, d \in A$. We call a function $u : A \to \mathbb{R}$ that satisfies (4) for all $a, b, c, d \in A$ 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.

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 3.7 below). We also characterise the larger class of BCPs with mixture-linear strong utilities. The latter characterisation is obtained by strengthening the SST condition in Theorem 3.3 to:

Axiom 7 (Strict Stochastic Transitivity [StST]) *For any $a, b, c \in A$,*

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

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 Axiom 6.*

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 and 2, and there exists a weak utility for P. It is therefore intuitive that the behavioural difference between the model in Theorem 3.4 and that in Theorem 3.5 should be characterised by the difference between SST and StST.

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*

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

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.

To ensure the continuity of G, we therefore 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 $d \in A$.

Theorem 3.6 *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 3.7 *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 3.7 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 StST and Solvability (Axioms 7 and 8). Strengthing SST to StST ensures that the function converting utility differences to choice probabilities is *strictly* increasing, while solvability ensures it is continuous.

4 Concluding remarks: experimental implications

As noted in the Introduction, our decomposition of CCI into Axioms 5 and 6 may be useful for experimental testing of models that embed EU in a Fechnerian noise structure. If CCI fails but Axiom 6 is supported by the data, then von Neumann and Morgenstern are rejected, but not Fechner.

Loomes and Sugden (1998, pp.594-5) themselves point the finger of blame at von Neuman and Morgenstern, but based on different reasoning. They detect what they call a *bottom edge effect*, which is inconsistent with any known stochastic form of EU. However, this conclusion is based on casual (albeit convincing) empiricism rather than formal testing. Our approach offers a formal avenue for confirming their conjecture. Indeed, when exactly one element of $\{a, b\}$ sits on the bottom edge of the Machina-Marschak triangle, then Axiom 6 rules out a particular type of bottom edge effect.

This testing strategy has the further advantage that both CCI and Axiom 6 can be tested on *aggregate* data; that is, on data which pools the choices of multiple subjects. This feature of CCI was exploited by Loomes and Sugden (1998) in their experiments. 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 properties of stochastic choice models are not robust to aggregation over heterogeneous individuals; aggregate choice proportions do not inherit these properties from individual choice probabilities.[5] This can seriously hamper the testing of such properties. In particular, experiments rarely confront the same individual with the same choice problem more than three times, so we usually have very limited data with which to identify individual choice probabilities. Axiom 5 suffers from precisely this problem, so the possibility of falsifying it *indirectly*, via tests of CCI and Axiom 6, may be an attractive proposition.

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-Aumann acts in a domain of uncertainty. On this domain, *subjective expected utility (SEU)* is mixture-linear. It follows that Axiom 6 is also the bridge from scalability with respect to an SEU "scale" to SEU maximisation with Fechnerian noise, at least within an Anscombe-Aumann (as opposed to Savage) environment.

[5] This is a variation on the Sonnenschein-Mantel-Debreu theme. It can also be seen as a version of Simpson's paradox.

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.

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.

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.

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.

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

Appendix

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 $\hat{u}$. □

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 binary relation $\succsim^P$ 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 $a, b, c \in A$

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

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

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

so $P(a,b) > \frac{1}{2}$ by SST, which contradicts $a \sim^P b$.

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

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

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

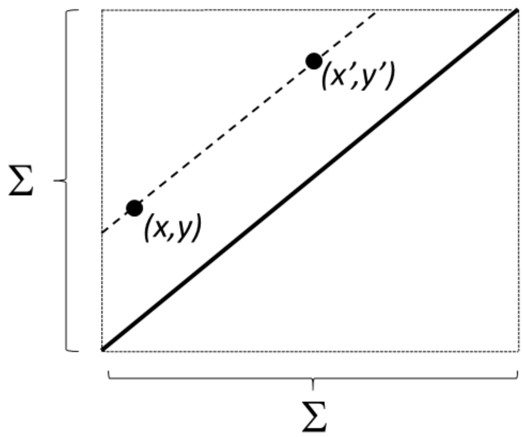

Figure 3: Domain of F

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 $\lambda \in (0, 1)$.[6] Let $a, b, a', b' \in A$ be such that $x = u(a)$, $y = u(b)$, $x' = u(a')$ and $y' = u(b')$. Then Axiom 6 and the mixture-linearity of u imply

$$F(x, y) = P(a, a\lambda b') = P(b'\lambda a, b') = F(x', y')$$

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. 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 StST, 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 $\succsim^P$, StST 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 7-8. The proof of Theorem 3.4 establishes that P is simply 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 strictly decreasing. This would

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

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 to obtain another violation of solvability.

Conversely, suppose P is simply scalable by (u, F) with u mixture-linear and F continuous in each argument. Theorem 3.4 ensures that P satisfies Axioms 1, 3, 5 and 7. It remains to verify Axiom 8. 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 $\lambda \in [0, 1]$. $\square$

Proof of Theorem 3.7. The result follows by the same argument as for Theorem 3.3, *mutatis mutandis.* $\square$