Available online at www.sciencedirect.com
SCIENCE DIRECT

Journal of Economic Theory 111 (2003) 172–191

JOURNAL OF Economic Theory

http://www.elsevier.com/locate/jet

Reference-dependent subjective expected utility

Robert Sugden*

School of Economic and Social Studies, University of East Anglia, Norwich NR4 7TJ, UK

Received 20 September 2000; final version received 20 May 2002

Abstract

A reference dependent generalisation of subjective expected utility theory is presented. In this theory, preferences between acts depend *both* on final outcomes *and* on reference points (which may be uncertain acts). It is characterised by a set of axioms in a Savage style framework. A restricted form of the theory separates attitudes to end states (encoded in a 'satisfaction function') from attitudes to gains and losses of satisfaction. Given weak additional assumptions, the restricted theory excludes cycles of choice, explains observed disparities between willingness to pay and willingness to accept valuations of lotteries, and predicts preference reversal.

JEL classification: D81

Keywords: Reference dependence; Loss aversion; Endowment Effect; Prospect theory; Preference reversal

1. Introduction

A large amount of evidence from experiments, from observations of markets and from contingent valuation studies suggests that decision-making behaviour often exhibits what has been variously called an endowment effect, loss aversion, or status quo bias.[1] In choice among lotteries, these effects show up in individuals' reluctance to undertake gambles which involve the possibility of loss, measured relative to current endowments—a reluctance that cannot be fully explained by risk aversion. In choice among commodity bundles, they show up in individuals' reluctance to accept

*Fax: +44 1603 250 434.
E mail address: r.sugden@uea.ac.uk.

[1] The modern literature on these effects stems mainly from [5,17]. There are reviews of the evidence in [1,16].

0022 0531/03/$ see front matter
doi:10.1016/S0022 0531(03)00082 6

trading opportunities that take them away from their current endowments. In contingent valuation studies, they show up as disparities between willingness-to-pay valuations and willingness-to-accept valuations of the same goods—disparities that are far too large to be explained by income and substitution effects. These effects can be interpreted as evidence that, contrary to the conventional assumptions of consumer theory and decision theory, individuals' preferences are *reference-dependent*—that is, preferences over final outcomes vary according to the reference point from which those outcomes are viewed. This paper will be concerned with reference-dependence in choice under uncertainty.

To date, the most fully developed reference-dependent theory of choice under uncertainty is *prospect theory*, proposed by Kahneman and Tversky [5]. In its original form, prospect theory applied only to a restricted class of simple lotteries, with given probabilities. Later developments have given the theory a rank-dependent form, extending its domain to all lotteries, and have allowed probabilities to be subjective [14,18]. However, even these more general forms of prospect theory have two significant limitations. The first is that preferences over lotteries depend only on the *changes* in wealth induced by those lotteries, and not on absolute *levels* of wealth. When applied to market behaviour, a theory of this kind will generate peculiar (and empirically false) implications. Suppose that, from some status quo position, an agent is willing to make a particular exchange. If she makes that exchange, she moves to a new status quo position. But then, according to the theory, she must be willing to make exactly the same exchange again; and so on indefinitely, until a boundary constraint is met. It seems clear that an adequate theory of reference-dependent preferences must take account of *both* changes in wealth *and* levels of wealth.[2]

The second limitation is that prospect theory does not apply to situations in which the agent's initial endowment is itself uncertain. Since the theory is constructed in a conceptual framework which does not include the notion of a state of the world, the probability distribution of gains and losses induced by a movement from one lottery to another is not well-defined. Thus, there is an important class of decision problems (including, for example, decisions about the sale of risky assets) to which prospect theory cannot be applied.

In the present paper, I propose a new theory of reference-dependent choice under uncertainty which overcomes these limitations. This theory is built within a Savage-style framework of states, consequences and acts, which does not presuppose any probability measure. Preferences are defined over acts, but are allowed to vary according to the reference point from which the relevant acts are viewed; the reference point can be any act, certain or uncertain.

In reference-dependent theories, the concept of 'reference point' is primitive, just as the concepts of 'consequence', 'state of the world' and 'preference' are in Savage's

[2] Kahneman and Tversky [5, pp. 277 278] note that 'strictly speaking', value should be defined as a function in two arguments changes in wealth relative to an initial asset position, *and* that position itself. However, they argue that the formulation they use in the paper 'generally provides a satisfactory approximation'.

theory. Thus, the interpretation of 'reference point' is not specified formally within the theory I propose. In discussing the implications of that theory, I shall follow a standard practice in the literature of reference-dependence, and interpret an agent's reference point as her current endowments. In allowing reference points to be uncertain acts, I do not depart from that standard interpretation, but merely take account of the fact that endowments can be state-dependent.[3]

The theory that I propose differs from conventional subjective expected utility theory *only* in respect of reference-dependence. In particular, it is linear in probabilities, and so cannot account for phenomena such as the Allais and Ellsberg paradoxes, which contravene Savage's sure-thing principle. Thus, the theory should not be interpreted as an attempt to explain all observed regularities in decision-making behaviour under uncertainty. It is presented as a possible model of how decision-making behaviour is affected by reference points; other complicating factors have been abstracted in the interests of conceptual clarity.

2. The subjected expected utility representation of reference-dependent preferences

I work within a theoretical framework that is broadly that of Savage [13]. There is a set S of *states*; an individual state is denoted by s. Any subset of S is an *event*; individual events are denoted by A, B, C; the power set of S, i.e. the collection of all events, is denoted by $\mathscr{S}$. My axioms will force $\mathscr{S}$ to be atomless. There is a set X of *consequences*; individual consequences are denoted by w, x, y, z. An act is a function from S to X; acts are denoted by f, g, h. An act is *simple* if it has a finite image. In this paper I consider only the set of simple acts, denoted by F.[4] Thus, all statements of the form 'for all f …' should be read as 'for all f in F'. An act f is *constant* if there is some x such that $f(s) = x$ for all s; this is written as $f = x$.

The main deviation from Savage's framework is in the definition of *preference* as a triadic rather than a binary relation. The preference relation is a subset of $F \times F \times F$; a typical proposition about preference is written as $f \succsim g|h$, and is read as 'f is weakly preferred to g, viewed from h', where h is a *reference act*; the relation itself will be denoted by $\succsim$. The preference relation is extended to consequences by means of preferences over constant acts, so that $x \succsim y|z$ is equivalent to '$f \succsim g|h$ where $f = x, g = y$ and $h = z$'; in such a case, z is the *reference consequence*. *Strict preference* and *indifference* between acts or consequences are defined from weak preference, so that $f \succ g|h \Leftrightarrow$ ($f \succsim g|h$ and not $g \succsim f|h$) and $f \sim g|h \Leftrightarrow$ ($f \succsim g|h$ and $g \succsim f|h$).

[3] I recognize that, in some problems to which reference dependent theory might be applied, there may be ambiguity about the precise definition of 'current endowments'. But some degree of openness in the interpretation of theoretical concepts is to be expected in any relatively new theory. If a theory proves useful, common understandings about the interpretation of its concepts tend to solidify as applications accumulate.

[4] The restriction to simple acts removes the need for Savage's P7 axiom in characterizing the expected utility representation [13, pp. 76–82]. It allows a similar simplification in the context of reference dependent preferences.

This triadic preference relation is used to represent the idea that preferences are reference-dependent. A decision problem can be described by a reference act (interpreted as the agent's status quo position) and an *opportunity set* of acts (the set of options from which the agent must choose), of which the reference act is one element. The agent chooses either to *stay* at the status quo or to *move* to one of the other options; this choice is determined by his preferences over the elements of the opportunity set, viewed from the reference act. Thus, the interpretation of $f \succeq g|h$ is that if the agent's status quo position is h, and if his opportunity set includes both f and g, then the option of moving to f is weakly preferred to the option of moving to g. Notice that the status quo position need not be a constant act.

Savage shows that, if a binary preference relation over acts satisfies certain postulates, it can be represented by a utility function and a probability measure: this is the *subjective expected utility representation* of binary preferences (for short, the *binary SEU representation*). My first objective is to generalize this result to reference-dependent preferences. I begin with some definitions.

Definition 1. A *probability measure* is a function $p: S \to [0,1]$ such that $p(\emptyset) = 0, p(S) = 1$, and for all disjoint $A, B: p(A) + p(B) = p(A \cup B)$.

Definition 2. Let p be any probability measure. Then E_p, the *expectation operator* for p, is defined so that, for any function $z: S \to \mathbb{R}$ with finite image, $E_p[z]$ is the mathematical expectation of z, calculated with respect to the probability measure p.

Definition 3. A *relative value function* is a finitely-valued function $v: X \times X \to \mathbb{R}$, such that for all $x: v(x,x) = 0$. An index of the form $v(x,z)$ is to be interpreted as a measure of the desirability of x, relative to z, when both are viewed from the reference consequence z.

Definition 4. A reference-dependent preference relation $\succeq$ has a *unique subjective expected utility representation* if there exists a unique probability measure p, and a relative value function v, unique up to multiplication by a positive constant, such that for all f, g, h:

$$f \succeq g|h \iff E_p[v(f,h) - v(g,h)] \geqslant 0. \tag{1}$$

This *reference-dependent SEU representation* is a natural generalization of Savage's binary SEU representation. The relative value function takes the place of Savage's utility function; (1) reduces to Savage's representation if, for all $x, z: v(x,z)$ is independent of z. Note also that it is an implication of (1) that, for any given reference consequence z, preferences *viewed from z* have a binary SEU representation in which the Savage utility function $U(.)$ takes the form $U(x) = v(x,z)$.

The essential ingredients for an SEU representation theorem can be found by adapting certain ideas from *general regret theory* [15]. In this theory, preferences over acts are defined relative to the agent's opportunity set. A typical preference proposition takes the form 'f is weakly preferred to g, given that the opportunity set is Z' (with $f, g \in Z$). This is written $f \succsim g/Z$. For cases in which the opportunity set has exactly three elements, $f \succsim g|h$ would serve as an alternative notation for $f \succsim g/\{f, g, h\}$. In this special sense, general regret theory postulates the existence of a triadic preference relation which ranks pairs of acts f, g in relation to some third act h; h serves as a point of reference for the comparison of f and g. Formally, the triadic preferences of general regret theory are similar to reference-dependent preferences. The restrictions that general regret theory imposes on triadic preferences are motivated by analogy with Savage's postulates. Since these analogies extend to the case of reference-dependent preferences, they provide the motivation for corresponding restrictions on such preferences. For brevity, I shall simply state these restrictions; their motivation is explained in [15].

First, more definitions are needed. The restriction of an act f to a non-empty event A is an *act component*, and is denoted by f_A. For each non-empty event A, a relation of *conditional preference* $\succsim_A$ is defined as follows:

Definition 5. For all f, g, h: $f \succsim_A g|h \Leftrightarrow$ [for all f', g', h': $(f'_A = f_A, g'_A = g_A, h'_A = h_A, f'_{S\backslash A} = g'_{S\backslash A}) \Leftrightarrow f' \succsim g'|h'$].

Propositions with the form $f \succsim_A g|h$ are read as 'f is weakly preferred to g, viewed from h, conditional on A'. Event A is *null* if $f \succsim_A g|h$ for all f, g, h. Act f is *constant in* A if there is some x such that $f(s) = x$ for all s in A. This is written as $f_A = x$; f_A is a *constant act component*. A binary relation $=_0$ on $\mathscr{S}$ is defined by:

Definition 6. A *transposition configuration* is an array $\langle A, B, f, g, h, f', g', h' \rangle$ such that A and B are non-null and disjoint, and (with C defined by $C \equiv S\backslash[A \cup B]$) there exist x, y, z, x', y', z' such that $x \succ y|z, x' \succ y'|z', f_A = f'_B = x, g_A = g'_B = y, h_A = h'_B = z, f_B = f'_A = y', g_B = g'_A = x', h_B = h'_A = z', f_C = f'_C, g_C = g'_C, h_C = h'_C$.

Definition 7. For non-null and disjoint $A, B: A =_0 B \Leftrightarrow$ [for all transposition configurations $\langle A, B, f, g, h, f', g', h' \rangle : f \succsim g|h \Leftrightarrow f' \succsim g'|h'$].

Note that each of the three acts f, g, h in Definition 6 is constant in A and in B. The acts f', g', h' are constructed from f, g, h by transposing consequences between A and B. If such a transposition induces a change in preference, this reveals an asymmetry (or difference in subjective 'weight') between A and B in their impact on preferences. Conversely, if indifference is preserved in such transpositions, A and B are revealed as having equal subjective weight. Thus, $=_0$ is the relation of *equal subjective probability*; propositions with the form $A =_0 B$ are to be read as 'A is exactly as probable as B'.

Now consider the following postulates:

R1. *Completeness*: For all f, g, h: $f \succsim g|h$ or $g \succsim f|h$.
R2. *Transitivity*: For all f, f', f'', h: $(f \succsim f'|h$ and $f' \succsim f''|h) \Rightarrow f \succsim f''|h$.
R3. *Existence of Conditional Preferences*: For all $A, f, g, h : [f_{S\backslash A} = g_{S\backslash A}$ and $f \succsim g|h] \Rightarrow f \succsim_A g|h$.

R4. *Sure Thing Principle*: For all f, g, h and for all disjoint A, B such that A is non-null: (i) $(f \succeq_A g|h$ and $f \succeq_B g|h) \Rightarrow f \succeq_{A \cup B} g|h$, and (ii) $(f \succ_A g|h$ and $f \succeq_B g|h) \Rightarrow f \succ_{A \cup B} g|h$.

R5. *Event Independence*: For all f, g, h, x, y, z and for all non-null A: $(f_A = x$ and $g_A = y$ and $h_A = z) \Rightarrow (f \succeq_A g|h \Leftrightarrow x \succeq y|z)$.

R6. *Existence of Qualitative Probability*: For all $A, B, f, g, h, f', g', h'$: $[\langle A, B, f, g, h, f', g', h' \rangle$ is a transposition configuration and $f \sim g|h$ and $f' \sim g'|h'] \Rightarrow A =_0 B$.

R7. *Non-triviality*: There exist x, y, z such that $x \succ y|z$.

R8. *State-space Continuity*: For all f, g, h such that $f \succ g|h$, and for any given x, y, z, there is some finite partition of S such that, for each event A in this partition, for all f', g', h': $[(f'_A = x$ or $f'_A = f_A)$ and $(g'_A = y$ or $g'_A = g_A)$ and $(h'_A = z$ or $h'_A = h_A)$ and $f'_{S \backslash A} = f_{S \backslash A}$ and $g'_{S \backslash A} = g_{S \backslash A}$ and $h'_{S \backslash A} = h_{S \backslash A}] \Rightarrow f' \succ g'|h'$.

R1–R8 are closely related to the postulates that Savage labels as P1–P6, and that he shows characterize the binary SEU representation. Completeness and Transitivity together extend Savage's ordering axiom P1 to the case of reference-dependent preferences. Similarly, Existence of Conditional Preferences, Event Independence, Non-triviality, and State-space Continuity are extensions of P2, P3, P5 and P6, respectively. The Sure Thing Principle extends the principle of the same name, as stated informally by Savage [13, pp. 21–26].[5] The only wholly new axiom among R1–R8 is Existence of Qualitative Probability, which substitutes for Savage's P4.[6]

The following theorem adapts Savage's representation theorem to the case of reference-dependent preferences (proofs of theorems, where not immediately obvious, are presented in the appendix):

Theorem 1. R1–R8 *are jointly equivalent to the existence of a unique reference-dependent SEU representation.*

[5] This principle plays an important part in the proof of Savage's representation theorem, but is not needed as an explicit axiom in Savage's system as it is implied by P1 and P2. This implication does not generalize to triadic preference relations.

[6] Savage's analysis of qualitative probability does not generalize straightforwardly to triadic preferences. The difficulties are related to the uses that Savage makes of the postulate that preferences over acts are transitive. In particular, he uses this postulate in deriving the transitivity of the relation 'is at least as probable as'. In a theory of transitive binary preferences, for any act f, if there is another act f' such that $f' \sim f$, we can infer that the ranking of f' with respect to any third act g is the same as that of f with respect to g. In other words, the fact that $f' \sim f$ allows us to *substitute* f' for f without affecting preferences. If preferences are reference dependent, such substitutions are illegitimate unless the reference act is held constant throughout. See [15] for a discussion of similar problems in the axiomatization of regret theory, in which there is a probability measure but preferences are non transitive.

3. A special case: satisfaction-change decomposability

In the remainder of the paper, I present a special case of the reference-dependent SEU representation. This allows the separation of two components of an agent's attitudes to choices among acts: 'attitudes to end states' and 'attitudes to gain and loss'.

From now on, I assume that R1–R8 hold, and hence that reference-dependent preferences have an SEU representation. I begin with the following set of definitions:

Definition 8. A *satisfaction function* is a function $u: X \to \mathbb{R}$.

Definition 9. A *gain/loss evaluation function* is defined in relation to a given satisfaction function $u(.)$. It is an increasing function $\varphi: R \to \mathbb{R}$, where $R = \{r \in \mathbb{R}\colon (\exists x, x' \in X) u(x) - u(x') = r\}$, with $\varphi(0) = 0$; $\varphi(r)$ is the *evaluation* of r.

Definition 10. A reference-dependent preference relation $\succeq$ has a *unique SEU representation with satisfaction-change decomposability* if there exist a unique probability measure $p(.)$, a satisfaction function $u(.)$, unique up to affine transformations, and a finitely-valued gain/loss evaluation function $\varphi(.)$, unique up to multiplication by a positive constant, such that, for all f, g, h:

$$f \succeq g | h \iff E_p[\varphi(u[f] - u[h]) - \varphi(u[g] - u[h])] \geqslant 0. \tag{2}$$

For short, I shall call this the *satisfaction-change SEU (or SCSEU) representation*. Note that (2) is equivalent to (1) with the added restriction that, for all x, z: $v(x, z) = \varphi(u[x] - u[z])$. If an agent's preferences are represented by (2), he acts as if maximising the mathematical expectation of the evaluation of changes in satisfaction. Thus, $u(.)$ might be interpreted as encoding the agent's attitudes to consequences in themselves: we might think of $u(x)$ as the psychological satisfaction to be expected from having x. Then $\varphi(.)$ might be interpreted as encoding the agent's attitudes to anticipated gains and losses of such satisfaction.

What restrictions on preferences allow such a decomposition? The necessary restrictions are of two kinds. First, it is necessary that preferences are such that *at least* one pair of $u(.)$ and $\varphi(.)$ exist, satisfying (2). Second, it is necessary that *only one* such pair exists (subject to the relevant normalizations). If this second condition is to be satisfied in general, we cannot continue to follow Savage's theoretical strategy of not imposing structure on X.[7] Accordingly, from now on I restrict attention to the

[7] As an example of a case in which X has insufficient structure, suppose that $X \quad \{x, y, z\}$. Suppose reference dependent preferences have an SEU representation with $v(x, z) \quad v(z, x) \quad 1$, $v(x, y) \quad v(y, x) \quad 0.4$, and $v(y, z) \quad v(z, y) \quad 0.8$. Let us construct corresponding $u(.)$ and $\varphi(.)$ functions, using the normalizations $u(z) \quad 0$, $u(x) \quad 1$, and $\varphi(1) \quad 1$. Defining $u(y) \quad a$, (2) is satisfied if and only if $\varphi(\ 1) \quad 1$, $\varphi(a) \quad \varphi(\ a) \quad 0.8$, $\varphi(1 \ a) \quad \varphi(a \ 1) \quad 0.4$, and $\varphi(.)$ must be an increasing function. All these conditions can be satisfied for any value of a in the interval $0.5 < a < 1$. Thus $u(.)$ and $\varphi(.)$ are not uniquely defined.

case in which $X = \mathbb{R}_+$. Consequences will be interpreted as levels of wealth. The imposition of this structure permits the following definition:

Definition 11. An SCSEU representation is *well-behaved* if $u(.)$ is continuous and increasing and if $\varphi(.)$ is continuous. (Recall that Definitions 9 and 10 already require $\varphi(.)$ to be increasing and finite-valued.)

Consider the following postulates:

S1. *Increasingness*: For all $x, y, z : x \geqslant y \Rightarrow x \succsim y|z$.

S2. *Consequence-space Continuity*: For all x, x', y, z: there exists some real number $\varepsilon > 0$ such that $(x - \varepsilon \leqslant x' \leqslant x + \varepsilon) \Rightarrow [(x \succ y|x \Rightarrow x' \succ y|z)$ and $(y \succ x|z \Rightarrow y \succ x'|z)$ and $(y \succ z|x \Rightarrow y \succ z|x')]$.

S3. *Gain/Loss Symmetry*: Let A, B be any events such that $A =_0 B$. Let w, x, y, z be any consequences such that $w > x$ and $y > z$. Let f, g, h, h' be any acts such that $f_A = h'_A = w, g_A = h_A = x, f_B = h_B = z, g_B = h'_B = y$. Then $f \succsim_{A \cup B} g|h \Leftrightarrow f \succsim_{A \cup B} g|h'$, and $f \succ_{A \cup B} g|h \Leftrightarrow f \succ_{A \cup B} g|h'$.

S4. *Gain/Loss Additivity*: Let A, B be any events such that $A =_0 B$. Let w, x, y, z be any consequences such that $w > x > y > z$ or $z > y > x > w$. Let f, g, h, f', g', h' be any acts such that $f_A = f'_A = w$, $g_A = g'_B = h_A = x, g'_A = h'_A = g_B = y, f_B = f'_B = h_B = h'_B = z$. Then $f \succsim_{A \cup B} g|h \Leftrightarrow f' \succsim_{A \cup B} g'|h'$, and $f \succ_{A \cup B} g|h \Leftrightarrow f' \succ_{A \cup B} g'|h'$.

Given the reference-dependent SEU representation, S1, S2, and S3 and S4 are respectively equivalent to the following restrictions on the relative value function:

S1.* For all x, z: $v(x, z)$ is increasing in x.

S2.* $v(x, z)$ is continuous in x and z.

S3.* For all w, x, y, z such that $w > x$ and $y > z$: $v(w, x) \gtreqless v(y, z) \Leftrightarrow v(x, w) \gtreqless v(z, y)$.

S4.* For all w, x, y, z such that $w > x > y > z$ or $z > y > x > w$: $v(w, x) \gtreqless v(y, z) \Leftrightarrow v(w, y) \gtreqless v(x, z)$.

Increasingness requires that larger consequences are preferred to smaller ones, irrespective of the reference point. Given that consequences are interpreted as levels of wealth, this restriction seems uncontroversial. Consequence-space Continuity is a technical condition, which supplements State-space Continuity. These two conditions, which make use of the structure that has been imposed on X, will be used in proving the uniqueness of $u(.)$ and $\varphi(.)$. The other two postulates require more explanation.

Given the postulates that characterize the reference-dependent SEU representation, and given any consequences w, x, y, z such that $w > x$ and $y > z$, we can meaningfully ask whether the desirability of moving from x to w is greater than, equal to, or less than the desirability of moving from z to y. The acts f, g, h, h', as defined in the statement of Gain/Loss Symmetry, constitute a test case. Recall that the interpretation of $A =_0 B$ is that A and B have equal subjective probability: other things being equal, consequences occurring in these two events have equal weight in determining preferences over acts. Thus, $f \succsim_{A \cup B} g|h$ signifies that the desirability of moving from x to w (which counts in favour of f in event A) is at least as great as the desirability of moving from z to y (which counts in favour of g in B). Similarly, $f \succsim_{A \cup B} g|h'$ signifies that moving from w to x is at least as *un*desirable as moving from y to z. Gain/Loss Symmetry imposes the restriction that either of these propositions about relative desirability implies the other.

Similar reasoning may be applied to the acts f, g, h, f', g', h', as defined in the statement of Gain/Loss Additivity. Consider the case in which $w > x > y > z$. (The case in which $z > y > x >$ w can be explained in a similar way, but with reference to the undesirability of losses rather than the desirability of gains.) Now, $f \succsim_{A \cup B} g|h$ signifies that moving from x to w is at least as desirable as moving from z to y. Similarly, $f' \succsim_{A \cup B} g'|h'$ signifies that moving from y to w is at least as desirable as moving from z to x. Gain/Loss Additivity imposes the restriction that either of these propositions implies the other.

The following representation theorem can be proved:

Theorem 2. *Given that* $X = \mathbb{R}_+$, R1–R8 *and* S1–S4 *are jointly equivalent to the existence of a unique well-behaved SCSEU representation of reference-dependent preferences.*

4. General properties of the satisfaction-change representation

Theorem 2 establishes that, if an agent's preferences satisfy the specified postulates, there exist $p(.)$, $u(.)$ and $\varphi(.)$ such that (2) holds; $p(.)$ is unique, $u(.)$ is unique up to affine transformations, and $\varphi(.)$ is unique up to multiplication by a positive constant. Thus, given sufficient information about an agent's reference-dependent preferences, it is possible to decompose his attitudes to lotteries into attitudes to events (encoded in $p[.]$), attitudes to consequences (encoded in $u[.]$), and attitudes to gain and loss (encoded in $\varphi[.]$).[8]

It is natural to ask how familiar properties of risk aversion and loss aversion, as observed in an agent's choices, relate to properties of $u(.)$ and $\varphi(.)$. Beginning with risk aversion, the first step is to adapt the conventional definition of risk aversion so that it applies to reference-dependent preferences. For a given SEU representation, I define a *subjectively certain* act as an act f which gives some outcome x with

[8] For such a decomposition to be possible in general, it is necessary to have information about the agent's preferences viewed from *non constant* reference acts. For example, consider the following alternative pairs of functional forms for $u(x)$ and $\varphi(r)$. In Case A, $u(x) \quad 1 \quad e^{-\beta x}$, where β (corresponding with the Arrow Pratt coefficient of absolute risk aversion) is a positive constant, and $\varphi(r) \quad r$. In Case B, $u(x) \quad x$ and $\varphi(r) \quad 1 \quad e^{-\beta r}$. Preferences over acts viewed from any *constant* reference act z are the same in each case, and are independent of z. (The relevant reference independent preferences have a binary SEU representation, in which the Savage utility function can be normalized to $U(x) \quad 1 \quad e^{-\beta x}$.) Thus, we cannot distinguish between these two distinct cases without considering non constant reference acts.

probability one. All other acts are *subjectively uncertain*. I define the following concept of risk aversion:

Definition 12. Given that $X = \mathbb{R}_+$, a reference-dependent preference relation $\succsim$ is *weakly* (respectively: *strictly*) *risk-averse* if, for all subjectively certain acts f, h, and for all subjectively uncertain acts g such that $E_p[g] = x$ where x is the consequence that f gives with probability one: $f \succsim g|h$ (respectively: $f \succ g|h$).

Thus, risk-aversion is the property that subjectively certain acts are preferred to actuarially equivalent subjectively uncertain acts, *viewed from subjectively certain acts*.[9]

Now consider how risk aversion, so defined, is encoded in the SCSEU representation. Recall that reference-dependent preferences over acts, viewed from any fixed reference consequence z, have a binary SEU representation in which $v(x, z)$ is the Savage utility function. Thus, whether the agent is risk-averse depends on whether $v(x, z)$ is concave in x. Given satisfaction-change decomposability, $v(x, z) = \varphi(u[x] - u[z])$. If reference-dependent preferences have a well-behaved SCSEU representation, then it is a sufficient condition for weak risk aversion that $u(.)$ and $\varphi(.)$ are both weakly concave. It is sufficient for strict risk aversion that one of $u(.)$ and $\varphi(.)$ is strictly concave and that the other is weakly concave.

Thus, in the context of the SCSEU representation, risk aversion is a composite phenomenon. It can result either from diminishing marginal satisfaction with respect to wealth, or from attitudes to gains and losses of satisfaction, or from a combination of the two.

I now consider attitudes to gain and loss. I first need a categorization of such attitudes which, like the categorization of attitudes to risk in Definition 12, refers directly to reference-dependent preferences. Since the term 'loss aversion' has come to be used to describe a property of the 'value function' in Tversky and Kahneman's [5,18], theory,[10] I introduce a new expression: *attitudes to exchange*. Intuitively, an agent is exchange-averse if, other things being equal, he prefers the status quo to other options in his opportunity set. In order to formalize the notion of 'other things being equal', we may consider cycles of potential exchanges, which if accepted would take the agent from an initial status quo position f_1, through one or more other status quo positions $f_2, \ldots, f_n$, back to f_1. An agent who chooses to go round such a cycle ends up with the same act as he started with, but engages in a number of acts of

[9] It might seem natural to think that it should be part of the definition of risk aversion that, viewed from *any* reference act, each subjectively certain act is preferred to all actuarially equivalent subjectively uncertain acts. But consider an agent who is endowed with a subjectively uncertain act f, and has the opportunity to exchange this for the actuarially equivalent subjectively certain act g. The preference $f \succ g|f$ does not necessarily indicate a love of risk taking: it may simply indicate aversion to moving away from the status quo.

[10] Kahneman and Tversky's value function is similar to a utility function, but it assigns real valued indices to *increments* of wealth. A value function $\upsilon(.)$ is loss averse if, for all increments of wealth $r > 0$: $\upsilon(r) < \quad \upsilon(\; r)$.

exchange. An agent who has a strict preference for such cycles is naturally classified as exchange-loving.[11] Hence the following definition:

Definition 13. A reference-dependent preference relation is *weakly exchange-averse* if, for all distinct acts $f_1, \dots, f_n$: $[f_2 \succsim f_1 | f_1$ and $\dots$ and $f_n \succsim f_{n-1} | f_{n-1}] \Rightarrow f_n \succsim f_1 | f_n$. It is *strictly exchange-averse* if, for all distinct acts $f_1, \dots, f_n$: $[f_2 \succsim f_1 | f_1$ and $\dots$ and $f_n \succsim f_{n-1} | f_{n-1}] \Rightarrow f_n \succ f_1 | f_n$.

It turns out that exchange aversion is closely related to the following property of the gain/loss evaluation function:

Definition 14. A gain/loss evaluation function $\varphi(.)$ with domain R has *weak zero-point concavity* if, for all $r', r'' \in R$ such that $r' < 0 < r''$: $\varphi(r')/r' \geqslant \varphi(r'')/r''$. If this inequality is strict, $\varphi(.)$ has *strict zero-point concavity*.

Notice that if $\varphi(.)$ is weakly (respectively: strictly) concave everywhere, then it necessarily satisfies weak (respectively: strict) zero-point concavity; but in general, the converse is not true. To gain some intuition about the concept of zero-point concavity, let A, B be disjoint events such that $p(A) = -r'/(r'' - r')$ and $p(B) = r''/(r'' - r')$. Let f, h be any acts such that f_A, f_B, h_A, h_B are constant act components, $u(f_A) - u(h_A) = r''$ and $u(f_B) - u(h_B) = r'$. Thus f and h give equal expected satisfaction. If preferences have a SCSEU representation, $h \succsim f | h$ is true if and only if $\varphi(r')/r' \geqslant \varphi(r'')/r''$. More generally, weak (respectively: strict) zero-point concavity of $\varphi(.)$ implies that, viewed from itself, every act is weakly (respectively: strictly) preferred to every other act which gives the same expected satisfaction.

If $\varphi(.)$ satisfies weak (respectively: strict) zero-point concavity, it is a necessary condition for the agent's being willing to move from the current status quo that the move results in a weak (respectively: strict) increase in expected satisfaction. Since satisfaction is not a reference-dependent concept, any sequence of such willing moves must result in a weak (respectively: strict) increase in expected satisfaction overall. Hence the following theorem:

Theorem 3. *If reference-dependent preferences have a well-behaved SCSEU representation, then weak (respectively: strict) zero-point concavity of $\varphi(.)$ implies that preferences are weakly (respectively: strictly) exchange-averse.*

One might ask whether zero-point concavity is necessary as well as sufficient for exchange aversion. Loosely, the answer is that exchange aversion forces $\varphi(.)$ to have zero-point concavity, except close to the upper and lower extremes of its domain. To be more precise, I define a payoff function as a function ζ (with a finite image) which assigns a finite real number (a payoff) to every state. (A payoff function is an uninterpreted formal object; the concept of 'payoff' is not to be interpreted as

[11] If certain continuity assumptions are made, such preferences are vulnerable to 'money pumps' in the sense that the agent is willing to pay for the privilege of making a cycle of exchanges.

equivalent to ‘consequence’.) Now consider any SCSEU representation. The preference relation $\succsim$ can be extended to the set of all payoff functions by requiring that, for all payoff functions ζ_1, ζ_2, ζ_3:

$$\zeta_1 \succsim \zeta_2 | \zeta_3 \;\Leftrightarrow\; E_p[\varphi(\zeta_1 - \zeta_3) - \varphi(\zeta_2 - \zeta_3)] \geqslant 0. \tag{3}$$

Thus, for each ζ_i, (3) treats each payoff $\zeta_i(s)$ as if it is a satisfaction index. If, for some ζ, every payoff is an element of $u(X)$, then corresponding to ζ there is an act $f \in F$ such that $u[f(s)] = \zeta(s)$ for all s; ζ can then be interpreted as a reduced form of f. However, this construction allows preferences to be defined over payoff functions for which no corresponding acts exist. The following theorem can be proved:

Theorem 4. *If reference-dependent preferences have a well-behaved SCSEU representation, and if $\varphi(.)$ does not satisfy weak (respectively: strict) zero-point concavity, then there exists a set of payoff functions over which preferences, as defined by (3), do not satisfy weak (respectively: strict) exchange aversion.*

Recall that a violation of zero-point concavity is a property of $\varphi(.)$ in relation to two real numbers, $r', r'' \in R$. For some such pairs of numbers r', r'', the proof that this property implies a violation of exchange aversion over payoff functions requires the construction of payoff functions whose payoffs lie outside the interval $[r', r'']$. If $u(.)$ is bounded, some of those payoffs may be outside the bounds of the satisfaction function, in which case there may be no violation of exchange aversion over acts. Nevertheless, if the aim is to explain why exchange aversion might be a general property of preferences, Theorem 4 suggests that there is little to be gained by trying to weaken zero-point concavity.

5. Implications for behaviour

I now briefly consider some of the empirical implications of SCSEU theory— that is, the theory of choice under uncertainty that assumes R1–R8 and S1–S4.

One of the merits of this theory is that reference acts need not be constant. This allows the theory to predict the behaviour of agents who, having been endowed with lotteries, are offered opportunities to sell them. In particular, it can explain the frequently observed disparity between willingness-to-pay and willingness-to-accept valuations of lotteries [6]. I begin with two definitions:

Definition 15. A *lottery* faced by an agent with initial wealth w is a function $\lambda : S \to \mathbb{R}$, satisfying the restriction that, for each state s, $w + \lambda(s) \geqslant 0$; $\lambda(s)$ is the *return* in state s. The act of *entering* λ is the act f such that for all s: $f(s) = w + \lambda(s)$.

Definition 16. Consider an agent with initial wealth w facing a lottery λ, and let f be the act of entering this lottery. The *willingness-to-accept* valuation of this lottery, denoted $\mathrm{WTA}(\lambda, w)$, is the increment of wealth such that $w + \mathrm{WTA}(\lambda, w) \sim f | f$. The *willingness-to-pay* valuation of the lottery, denoted $\mathrm{WTP}(\lambda, w)$, is the increment of wealth such that $f \sim w + \mathrm{WTP}(\lambda, w) | w + \mathrm{WTP}(\lambda, w)$.

If preferences have a well-behaved SCSEU representation, $\mathrm{WTA}(\lambda, w)$ and $\mathrm{WTP}(\lambda, w)$ are uniquely defined for any given λ and w. Note that these two valuations are defined in relation to one another so as to screen out wealth effects. That is, if preferences are independent of reference acts, $\mathrm{WTA}(\lambda, w) \equiv \mathrm{WTP}(\lambda, w)$. However, if preferences are strictly exchange averse, $f \sim w + \mathrm{WTP}(\lambda, w) | w + \mathrm{WTP}(\lambda, w)$ and $w + \mathrm{WTA}(\lambda, w) \sim f | f$ jointly imply $w + \mathrm{WTA}(\lambda, w) \succ w + \mathrm{WTP}(\lambda, w) | w + \mathrm{WTA}(\lambda, w)$. Because of Increasingness, this implies $\mathrm{WTA}(\lambda, w) > \mathrm{WTP}(\lambda, w)$. Hence, by virtue of Theorem 3:

Theorem 5. *For all lotteries λ and for all initial wealth levels w: if preferences have a well-behaved SCSEU representation, strict zero-point concavity of $\varphi(.)$ implies* $\mathrm{WTA}(\lambda, w) > \mathrm{WTP}(\lambda, w)$.

SCSEU theory may also contribute to the explanation of *preference reversal*. This phenomenon, first discovered by Lichtenstein and Slovic [8] and Lindman [9], has been replicated in many different experimental designs. It is a discrepancy between responses to three experimental tasks. These tasks involve two lotteries. One, the '\$ bet', offers a relatively large return with relatively low probability; the other, the 'P bet', offers a smaller return with a higher probability. In the *choice* task, a subject is given no endowment and is offered a choice between the lotteries. In each of two *valuation* tasks, the subject is endowed with one or other of the lotteries and reports her WTA valuation of it. Apparent preferences between the two lotteries, as indirectly revealed in the valuation tasks, are systematically different from those directly revealed in the choice task: the \$ bet is more likely to be preferred in the valuation tasks than in the choice task.

In the perspective of a theory of reference-dependent preferences, such a 'reversal' need not be interpreted as an inconsistency. In the choice task, the subject reports her preferences as viewed from a reference point in which she owns neither lottery. In each valuation task, she reports her preferences as viewed from a reference point in which she owns one of the lotteries. Because the reference points differ between the tasks, the subject's ranking of acts need not be consistent across those tasks. Since the \$ bet is 'less similar to' certain money than is the P bet, we might expect a given propensity to exchange aversion to have a greater (positive) impact on the WTA valuation of the \$ bet than on that of the P bet.

I now firm up this intuition by showing that SCSEU theory predicts preference reversal under the special assumption that $\varphi(.)$ is strictly concave (an assumption that is compatible with, but stronger than, exchange aversion). The formal result is the following:

Theorem 6. *Assume that preferences have a well-behaved SCSEU representation. Let A', A'' be any events such that $1 > p(A'') > p(A') > 0$. Let c', c'' be real numbers such* *that $c' > c'' > 0$. Let λ', λ'' be lotteries defined by $\lambda'(A') = c'$, $\lambda'(S \backslash A') = 0$, $\lambda''(A'') = c''$, $\lambda''(S \backslash A'') = 0$. Let f', f'' be the acts of entering these lotteries, and let w be initial wealth. Suppose that $f' \sim f'' | w$. Then, if $\varphi(.)$ is strictly concave, $\mathrm{WTA}(\lambda', w) > \mathrm{WTA}(\lambda'', w)$.*

We may interpret λ' as a \$ bet and λ'' as a P bet. The condition $f' \sim f'' | w$ can be interpreted as indifference between the two bets, viewed from a reference point at which the agent has her initial wealth but owns neither bet (i.e. the reference point for the choice task). Given the assumption that $\varphi(.)$ is strictly concave, Theorem 7 implies that an individual who is indifferent between the two bets in the choice task will give a higher WTA valuation to the \$ bet than to the P bet.[12]

Much of the experimental and survey evidence of WTA/WTP disparities and of preference reversal involves lotteries whose payoffs are tiny relative to the liquid assets (let alone lifetime wealth) of a typical subject. It seems that very small variations in an agent's reference point can lead to significant changes in preferences over acts in the neighbourhood of that point. Considering any theoretical explanation of such small-scale effects, it is appropriate to ask whether the assumptions being used have credible implications for larger-scale decisions. As Rabin [12] shows, the claim that conventional expected utility theory can explain the risk aversion of experimental subjects is vulnerable to just such a criticism: that explanation comes at the cost of incredible implications about risk aversion over larger wealth intervals. But one of the virtues of SCSEU theory is that it *is* compatible with exchange aversion, as observed both in the small and in the large. The scale on which gains and losses of satisfaction are measured has a natural zero (representing the categorical distinction between loss and gain). It is both theoretically legitimate and psychologically defensible[13] to assume a discontinuity in $\varphi'(.)$ at this point (i.e. that the marginal evaluation of gains is less than the marginal evaluation of losses). If such a discontinuity is assumed, exchange aversion will be observed even with respect to infinitesimal gains and losses.

For the same reasons, SCSEU theory can explain risk aversion in the small, as observed for experimental subjects, without falling foul of Rabin's critique. For example, several recent experiments have found that, if subjects face a large number of choices between pairs of lotteries, their responses converge towards a pattern that can be represented by a stochastic form of expected utility theory; utility functions

[12] It is crucial for this explanation of preference reversal that valuation tasks elicit WTA valuations. Given the assumptions invoked by Theorem 6, SCSEU theory does not generate clear predictions for preference reversal experiments which elicit WTP valuations. In fact, there have been very few experiments of this latter type. It seems that this treatment tends to reduce the frequency of preference reversals in the normal direction [7,8] and to induce reversals in the opposite direction, i.e. reversals in which the \$ bet is chosen in the choice tasks but the P bet has the higher valuation [2]. These findings are consistent with the hypothesis that exchange aversion is at least a contributory cause of preference reversal.

[13] Kahneman and Tversky [5] justify their assumption of 'loss aversion' on psychological principles. This assumption implies a similar discontinuity in the slope of the value function at its natural zero (see footnote 10).

have the form $U = Y^{1-\alpha}$, where U is utility, Y is monetary gain in the experiment, and α (the Arrow–Pratt measure of relative risk aversion) is a constant satisfying $0<\alpha<1$. The estimated value of α can be as high as 0.7.[14] This pattern is not easily reconciled with expected utility theory in its conventional form, in which utility depends only on *levels* of wealth. The difficulty is that constant relative risk aversion for *changes* in wealth implies disproportionately large variations in absolute risk aversion in the particular neighbourhood of the agent's current wealth. From the perspective of conventional expected utility theory, this property of the utility function has to be interpreted as sheer coincidence. In contrast, the observed pattern of behaviour can be represented straightforwardly in SCSEU theory by assuming $u(.)$ to be approximately linear over the relevant range of wealth and setting $\varphi(r) = r^{1-\alpha}$ for $r \geqslant 0$.

6. Conclusion

The existence of endowment effects is now well established. The conclusion seems inescapable that individuals' preferences are not, as is assumed in most decision theory, independent of their status quo positions. I have proposed a theory of choice under uncertainty in which preferences are reference-dependent. In its most general form, this theory is a generalization of Savage's subjective expected utility theory. Thus, I have shown that insights about the significance of reference points, deriving from cognitive psychology and behavioural economics, can be integrated into decision theory at its most fundamental level. I have also proposed a restricted form of the theory, in which the subjective expected utility representation has a simple and tractable functional form. This version of the theory can explain some systematic features of observed decision-making behaviour, which are inconsistent with conventional theory. I hope that these achievements will stimulate further work in what remains a puzzlingly under-researched area of decision theory.

Acknowledgments

This paper was written as part of the Risk and Human Behaviour Programme of the Economic and Social Research Council of the UK (Award No. L 211 252 053). This work was also supported by the Leverhulme Trust. It has grown out of discussions with Ian Bateman, Robin Cubitt, John Hey, Graham Loomes, Daniel Kahneman, Alistair Munro, Prasanta Pattanaik and Chris Starmer. I am grateful for comments from an associate editor and an anonymous referee.

[14] These experiments are described by Hey and Orme [4], Loomes and Sugden [11], and Hey [3]. The findings reported in the text derive from new panel data analysis of these experiments [10].

Appendix

Proof of Theorem 1. I begin by proving that R1–R8 imply the reference-dependent SEU representation. The proof exploits isomorphisms between reference-dependent preferences and the preferences postulated by general regret theory. □

Lemma 1. *Completeness, Existence of Conditional Preferences, Sure Thing Principle, Event Independence, Existence of Qualitative Probability, Non-triviality, and State-space Continuity jointly imply the existence of a unique probability measure* $p(.)$, *and a finite-valued function* $\psi : X \times X \times X \to \mathbb{R}$, *skew-symmetric in its first two arguments (i.e. for all* x, y, z: $\psi[x, y, z] = -\psi[y, x, z]$) *and unique up to multiplication by a positive constant, such that for all* f, g, h:

$$f \succsim g|h \Leftrightarrow E_p[\psi(f_E, g_E, h_E)] \geqslant 0. \tag{A.1}$$

Proof. The representation theorem for general regret theory [15, Theorem 2] uses seven axioms, labelled Q1(i) and Q2–Q7. These axioms impose restrictions on preferences of the form $f \succsim g/Z$, where Z is the opportunity set (with $f, g \in Z$). If we write $f \succsim g/\{f, g, h\}$ as $f \succsim g|h$, these axioms impose corresponding restrictions on reference-dependent preferences. The restrictions imposed on $\succsim$ by Completeness, Existence of Conditional Preferences, Sure Thing Principle, Event Independence, Existence of Qualitative Probability, and State-space Continuity respectively are either equivalent to, or stronger than, those imposed by Q1(i), Q2, Q3, Q4, Q5 and Q7.[15] Q6 requires that there exist x, y such that $x \succsim y/\{x, y\}$. Clearly, this axiom does not apply to three-element feasible sets. However, the only role of Q6 in the proof of the representation theorem is to ensure the existence of *some* strict preference. The proof works just as well if the restriction is that there exist x, y, z such that $x \succsim y/\{x, y, z\}$, which is equivalent to the restriction imposed by Non-triviality. If the representation theorem is revised in this way, Lemma 1 is a corollary of it.[16] □

Notice that Transitivity is not used in this representation theorem. The next lemma shows the effects of imposing Transitivity in addition to the other axioms:

Lemma 2. *If Completeness, Existence of Conditional Preferences, Sure Thing Principle, Event Independence, Existence of Qualitative Probability, Non-triviality, and State-space Continuity all hold, and if reference-dependent preferences are represented by* (A1), *then Transitivity implies that for all consequences* w, x, y, z: $\psi(w, x, z) + \psi(x, y, z) + \psi(y, w, z) = 0$.

[15] Readers who consult [15] should note that there is an error in the statement of Q5. The error is equivalent to omitting the condition $h_C \quad h'_C$ in Definition 2.

[16] The formal statement of the theorem in [15] does not include the claim that $\psi(., ., .)$ is a finite valued function, but the method by which this function is constructed in the theorem forces it to take finite values.

Proof. Suppose there exist w, x, y, z such that $\psi(w, x, z) + \psi(x, y, z) + \psi(y, w, z) > 0$. Let $\{A, B, C\}$ be a partition of S such that $p(A) = p(B) = p(C) = 1/3$. (Since $\mathscr{S}$ is atomless, such a partition must exist.) Define acts f, f', f'', h by $f_A = f'_B = f''_C = w$, $f_B = f'_C = f''_A = x$, $f_C = f'_A = f''_B = y$, $h = z$. Then (A1) implies $f' \succ f|h, f'' \succ f'|h, f \succ f''|h$, contrary to Transitivity. Suppose instead that $\psi(w, x, z) + \psi(x, y, z) + \psi(y, w, z) < 0$. If f, f', f'' are defined as before, (A1) implies $f \succ f'|h, f' \succ f''|h, f'' \succ f|h$, contrary to Transitivity. □

Now assume that all of R1–R8 hold. Then, by Lemma 1, there exist functions $p(.)$ and $\psi(.,.,.)$ such that preferences are represented by (A1). Define a function v: $X \times X \rightarrow \mathbb{R}$ by setting $v(x, y) = \psi(x, y, y)$ for all x, y. Because of the skew-symmetry property of ψ, $v(x, x) = 0$ for all x. It follows from the definition of $v(.,.)$ and from the skew-symmetry property that, for all x, y, z: $v(x, z) - v(y, z) = \psi(x, z, z) + \psi(z, y, z)$. By Lemma 2 and skew-symmetry, $\psi(x, z, z) + \psi(z, y, z) = \psi(x, y, z)$. Substituting $v(x, z) - v(y, z) = \psi(x, y, z)$ into (A1) for all x, y, z gives (1), the SEU representation. The only transformations of $v(.,.)$ that preserve preferences in this representation are multiplications by positive constants. This completes the proof that R1–R8 imply the reference-dependent SEU representation.

I now outline the proof of the converse. It is straightforward to show that Completeness, Transitivity, Existence of Conditional Preferences, Sure Thing Principle, and Event Independence are implied by the reference-dependent SEU representation. The proof that Existence of Qualitative Probability is implied by this representation works by showing that if $\langle A, B, f, g, h, f' g', h' \rangle$ is a transposition configuration, the conjunction of $f \sim g|h$ and $f' \sim g'|h'$ implies $p(A) = p(B)$, which in turn implies $A =_0 B$. That Non-triviality is implied by the representation follows from the requirement that $p(.)$ is unique. If Non-triviality does not hold, every act is indifferent to every other, viewed from every reference act, which implies $u(x, z) = 0$ for all x, z. But then preferences are independent of the probabilities assigned to events. That State-space Continuity is implied by the representation follows from the fact that $v(.,.)$ is required to be finite-valued, while $\mathscr{S}$ is atomless. □

Proof of Theorem 2. First, I prove that R1–R8 and S1–S4 imply that preferences can be represented by (2). Assume that $X = \mathbb{R}_+$, and that R1–R8 and S1–S4 hold. Since Theorem 1 establishes that R1–R8 imply the SEU representation, we can take as given a relative value function $v(.,.)$, unique up to multiplication by a positive constant. Because of Increasingness and Consequence-space Continuity, $v(.,.)$ is increasing in its first argument and continuous in both arguments. It is sufficient to show that, given an arbitrary normalization of $v(.,.)$, there exist a continuous and increasing satisfaction function $u(.)$, unique up to affine transformations, and a gain/loss evaluation function $\varphi(.)$, unique up to multiplication by a positive constant, such that $v(x, z) = \varphi[u(x) - u(z)]$ for all x, z. (Note that, for *given* normalizations of $v(.,.)$ and $u(.)$, $\varphi(.)$ is uniquely defined. However, permissible transformations of $v(.,.)$ and $u(.)$ imply corresponding transformations of $\varphi(.)$.)

Define $x_0 = 0$ and take any $x_1 > 0$. Fix a normalization of $v(.,.)$. By Increasingness, $v(x_1, x_0) > 0$. Using Increasingness and Consequence-space Continuity, *either* there exists a unique $x_2 > x_1$ such that $v(x_2, x_1) = v(x_1, x_0)$, *or* for all $x > x_1, v(x, x_1) < v(x_1, x_0)$. If such an x_2 exists, then *either* there exists a unique $x_3 > x_2$ such that $v(x_3, x_2) = v(x_2, x_1)$, *or* for all $x > x_2$, $v(x, x_2) < v(x_2, x_1)$; and so on. Thus, exactly one of the following two cases must be true.

Case 1: There exists a unique finite sequence $\langle x_0, x_1, x_2, \ldots, x_m \rangle$ such that, for all i in the interval $2 \leqslant i \leqslant m$, $v(x_i, x_{i-1}) = v(x_{i-1}, x_{i-2})$, and for all $x > x_m$, $v(x, x_m) < v(x_m, x_{m-1})$.

Case 2: There exists a unique infinite sequence $\langle x_0, x_1, x_2, \ldots \rangle$ such that, for all $i \geqslant 2, v(x_i, x_{i-1}) = v(x_{i-1}, x_{i-2})$.

Without making any assumption about which case holds, let X' be the set of elements in the relevant sequence.

Now consider any $i \geqslant 0$ and $k > 1$ such that $x_{i+k+1} \in X'$. By construction, $v(x_{i+k+1},\ x_{i+k}) = v(x_{i+1},\ x_i) = v(x_1, x_0)$. Using S4* (which is an implication of Gain/Loss Additivity), $v(x_{i+k+1}, x_{i+1}) = v(x_{i+k},\ x_i)$. Repeated application of this result establishes that, for all $x_i,\ x_j \in X'$ such that $i \geqslant j$: $v(x_i, x_j)$ depends only on (and is increasing in) $i - j$. Because of Gain/Loss Symmetry, it is also the case that, for all $i \geqslant 0$ and $k > 0$ such that $x_{i+k+1} \in X'$: $v(x_i, x_{i+1}) = v(x_{i+k},\ x_{i+k+1}) = v(x_0,\ x_1)$. By a similar use of S4*, it can be shown that, for all $x_i,\ x_j \in X'$ such that $i \leqslant j$: $v(x_i,\ x_j)$ depends only on (and is increasing in) $i - j$. Thus, there exists an increasing function $\rho(.)$ such that, for all $x_i,\ x_j \in X'$: $v(x_i,\ x_j) = \rho(i - j)$.

Now define an increasing function u^*: $X' \to \mathbb{R}_+$ by the condition that, for all $x_i \in X'$: $u^*(x_i) = iv(x_1,\ x_0)$. Define R' as the set of all values of $u^*(x_i) - u^*(x_j)$, i.e. $R' = \{r \in \mathbb{R}\colon (\exists x_i,\ x_j \in X')(i - j)v(x_1,\ x_0) = r\}$. Finally, define an increasing function $\varphi^* : R' \to \mathbb{R}$ by the condition that $\varphi^*([i - j]v[x_1,\ x_0]) = \rho(i - j)$; notice that $\varphi^*(0) = 0$. This construction guarantees that, for all $x_i,\ x_j \in X'$: $v(x_i, x_j) = \varphi^*[u^*(x_i) - u^*(x_j)]$. Since $v(.,.)$ is finitely-valued, so too is $\varphi^*(.)$.

Next, I show that $u^*(.)$ and $\varphi^*(.)$ (and their permissible transformations) are the *only* satisfaction and gain/loss evaluation functions that can satisfy $v(x_i,\ x_j) = \varphi[u(x_i) - u(x_j)]$ with respect to consequences in X'. Notice that, once we have set $x_0 = 0$ and have fixed the value of x_1, the subsequent elements $x_2,\ x_3, \cdots$ of the sequence are uniquely determined by $v(.,.)$, independently of how that function is normalized. Since the gain/loss evaluation function is increasing, (2) can be satisfied only if, for all $x_i \in X'$ satisfying $i > 0$: $u(x_i) - u(x_{i-1}) = u(x_1) - u(x_0)$. Thus it is a necessary condition for a satisfaction function to be compatible with (2) that its restriction to X' is an affine transformation of $u^*(.)$. For given normalizations of $v(.)$ and $u(.)$, $\varphi(.)$ is uniquely determined by the requirement that $\varphi[u(x_i) - u(x_j)] = v(x_i,\ x_j)$. Thus, it is a necessary condition for a gain/loss evaluation function to be compatible with (2) that its restriction to X' can be derived from $\varphi^*(.)$ by multiplication by a positive constant.

As so far described, this construction uniquely identifies the restriction of $u(.)$ to X'. But by setting the value of x_1 sufficiently close to zero, we can ensure that the number of elements of X' in any given finite interval of X is arbitrarily large. Because of the continuity properties of $v(.,.), u(.)$ is a continuous function. Thus, by taking the limit as $x_1 \to 0$, this construction identifies $u(.)$ up to affine transformations. Given any normalization of $u(.), \varphi(.)$ is uniquely determined by $v(.,.)$. Because $v(.,.)$ and $u(.)$ are continuous functions, so too is $\varphi(.)$.

It remains only to show that the representation (2) implies R1–R8 and S1–S4. Since (2) is a special case of the SEU representation, it follows immediately from Theorem 1 that R1–R8 are satisfied. Given that $u(.)$ and $\varphi(.)$ are continuous and increasing, it is straightforward to show that (2) satisfies S1–S4. □

Proof of Theorem 4. Assume that preferences have a well-behaved SCSEU representation, and suppose that $\varphi(.)$ fails to satisfy weak zero-point concavity. Since $\varphi(.)$ is continuous, there exist rational numbers r', $r'' \in R$ such that $r' < 0 < r''$ and $\varphi(r')/r' < \varphi(r'')/r''$. Since r', r'' are rational numbers, we can pick two positive integers n', n'' such that $n'/n'' = -r'/r''$. Now consider an $(n' + n'') \times (n' + n'')$ matrix of real numbers. Let b_{ij} be the element in the ith row and the jth column. For all i, j: $b_{ij} \in \{r', r''\}$. For each row i, there are exactly n' columns j for which $b_{ij} = r''$. For each column j, there are exactly n'' rows i for which $b_{ij} = r'$. Such a matrix (a variant of a Latin square) can be constructed for any n', n''. Let $E_j (j = 1, \ldots, n' + n'')$ be equally probable events. Let ζ_i $(i = 1, \ldots, n' + n'')$ be payoff functions. For all j, and for all $s \in E_j$, define $\zeta_1(s) = b_{1j}$ and $\zeta_i(s) = b_{ij} + b_{i-1,j}$ $(i = 2, \ldots, n' + n'')$. By construction, $\zeta_{n'+n''}(s) = 0$ for all s. Using (3), $r''\varphi(r') - r'\varphi(r'') > 0$ implies $\zeta_1 \succ \zeta_{n'+n''}$, and $\zeta_i \succ \zeta_{i-1}$ for $i = 2, \ldots, n' + n''$. Thus, preferences over this set of payoff functions violate weak exchange aversion. The proof for the case of strict zero-point concavity follows the same strategy. □

Proof of Theorem 6. Assume a well-behaved SCSEU representation in which $\varphi(.)$ is strictly concave. Let A be any event such that $1 > p(A) > 0$. Let λ be the lottery defined by $\lambda(A) = c$ (where $c > 0$) and λ $(S \backslash A) = 0$. Let the agent's initial wealth be w. Define $\pi \equiv p(A)$, $r \equiv u(w + c) - u(w)$, $t \equiv u(w + \mathrm{WTA}[\lambda, w]) - u(w)$; notice that $r > t > 0$. Using Definition 16, $\pi\varphi(t - r) + (1 - \pi)\varphi(t) = 0$. Now, holding w constant, let π and c vary together in such a way that the value of $\pi\varphi(r)$ remains constant, and consider the value of dt/dr. Notice that this procedure defines a set of lotteries, each of which has the same form as λ, such that the acts of entering those lotteries are mutually indifferent, viewed from the reference consequence w. Thus λ' and λ'', as defined in the statement of the theorem, belong to one such set. The theorem can be proved by showing that, within such a set, increases in c are always associated with increases in $\mathrm{WTA}(\lambda, w)$. Because $u(.)$ is an increasing function, it is sufficient to show that $dt/dr > 0$. In fact,

$$\begin{aligned} dt/dr = \pi[\varphi'(t - r) - \{\varphi(t) - \varphi(t - r)\}\varphi'(r)/\varphi(r)]/ \\ [\pi\varphi''(t - r) + (1 - \pi)\varphi'(t)]. \end{aligned} \tag{A.2}$$

Since $\varphi(.)$ is increasing, the denominator of this expression is strictly positive. If $\varphi(.)$ is strictly concave, the numerator is strictly positive too. Hence $dt/dr > 0$. □

References

[1] C. Camerer, Individual decision making, in: J.H. Kagel, A.E. Roth (Eds.), Handbook of Experimental Economics, Princeton University Press, Princeton, 1995, pp. 587 703.
[2] J.T. Casey, Reversal of the preference reversal phenomenon, Organ. Behav. Hum. Decision Processes 48 (1991) 224 251.
[3] J.D. Hey, Does repetition improve consistency?, Exp. Econom. 4 (2001) 5 54.
[4] J.D. Hey, C. Orme, Investigating generalizations of expected utility theory using experimental data, Econometrica 62 (1994) 1291 1326.
[5] D. Kahneman, A. Tversky, Prospect theory: an analysis of decision under risk, Econometrica 47 (1979) 263 291.
[6] J.L. Knetsch, J.A. Sinden, Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of value, Quart. J. Econom. 99 (1984) 507 521.
[7] M. Knez, V. Smith, Hypothetical valuations and preference reversals in the context of asset trading, in: A.E. Roth (Ed.), Laboratory Experimentation in Economics: Six Points of View, Cambridge University Press, Cambridge, pp. 131 154.
[8] S. Lichtenstein, P. Slovic, Reversals of preferences between bids and choices in gambling decisions, J. Exp. Psychol. 89 (1971) 46 55.
[9] H. Lindman, Inconsistent preferences among gambles, J. Exp. Psychol. 89 (1971) 390 397.
[10] G. Loomes, P. Moffatt, R. Sugden, A microeconometric test of alternative stochastic theories of risky choice, J. Risk Uncertainty 24 (2002) 103 130.
[11] G. Loomes, R. Sugden, Testing different stochastic specifications of risky choice, Economica 65 (1998) 581 598.
[12] M. Rabin, Risk aversion and expected utility theory: a calibration theorem, Econometrica 68 (2000) 1281 1292.
[13] L. Savage, The Foundations of Statistics, Wiley, New York, 1954.
[14] C. Starmer, R. Sugden, Violations of the independence axiom in common ratio problems: an experimental test of some competing hypotheses, Ann. Oper. Res. 19 (1989) 79 102.
[15] R. Sugden, An axiomatic foundation for regret theory, J. Econom. Theory 60 (1993) 159 180.
[16] R. Sugden, Alternatives to the neo classical theory of choice, in: I.J. Bateman, K.G. Willis (Eds.), Valuing Environmental Preferences, Oxford University Press, Oxford, 1999, pp. 152 180.
[17] R. Thaler, Toward a positive theory of consumer choice, J. Econom. Behav. Organ. 1 (1980) 39 60.
[18] A. Tversky, D. Kahneman, Advances in prospect theory: cumulative representation of uncertainty, J. Risk Uncertainty 5 (1992) 297 323.