The arguments of utility: Preference reversals in expected utility of income models

Lindsay, Luke

Abstract: There is a debate in the literature about the arguments of utility in expected utility theory. Some implicitly assume utility is defined on final wealth whereas others argue it may be defined on initial wealth and income separately. I argue that making income and wealth separate arguments of utility has important implications that may not be widely recognized. A framework is presented that allows the unified treatment of expected utility models and anomalies. I show that expected utility of income models can predict framing induced preference reversals, a willingness to pay-willingness to accept gap for lotteries, and choice-value preference reversals. The main contribution is a theorem. It is proved that for all utility functions where initial wealth and income enter separately, either there will be preference reversals or preferences can be represented by a utility function defined on final wealth alone

DOI: https://doi.org/10.1007/s11166-013-9162-z

J Risk Uncertain (2013) 46:175–189
DOI 10.1007/s11166-013-9162-z

The arguments of utility: Preference reversals in expected utility of income models

Luke Lindsay

Published online: 21 March 2013

Abstract There is a debate in the literature about the arguments of utility in expected utility theory. Some implicitly assume utility is defined on final wealth whereas others argue it may be defined on initial wealth and income separately. I argue that making income and wealth separate arguments of utility has important implications that may not be widely recognized. A framework is presented that allows the unified treatment of expected utility models and anomalies. I show that expected utility of income models can predict framing induced preference reversals, a willingness to pay-willingness to accept gap for lotteries, and choice-value preference reversals. The main contribution is a theorem. It is proved that for all utility functions where initial wealth and income enter separately, either there will be preference reversals or preferences can be represented by a utility function defined on final wealth alone.

Keywords Expected utility theory · Risk aversion · Preference reversals

JEL Classification C90 · D81

Expected utility theory (EUT) has remained the standard theory of choice used in economics. One of the appealing features of the theory is that risk aversion can be modeled using a concave utility function. Rabin (2000), however, has argued that a concave utility function cannot accommodate risk aversion over modest stakes without predicting implausible degrees of risk aversion over large stakes. If we accept this

L. Lindsay (✉)
Department of Economics, University of Zurich, Blümlisalpstrasse 10, 8006 Zurich, Switzerland
e-mail: luke.lindsay@econ.uzh.ch

argument, risk aversion over small stakes is an anomaly for the theory. This paper aims to clarify the implications of Rabin's paradox and different ways of defining utility in EUT.

The implications of Rabin's paradox have been debated in the literature. The paradox depends on the assumption that utility is defined on wealth but von Neumann and Morgenstern (1947) did not specify the domain of the utility function when they developed EUT. Cox and Sadiraj (2006) argue that utility can be defined on wealth or income and show that an expected utility of income model can explain risk aversion over small and large stakes.[1] Accordingly, they write: "the type of global small-stakes risk aversion assumed in previous literature (Rabin 2000; Rabin and Thaler 2001) has no implication for the expected utility of income model, hence no general implication for expected utility theory". Rubinstein (2006) makes a similar argument, also pointing out that Rabin's Paradox can be avoided by allowing utility to be defined on changes in wealth.[2] In contrast, Wakker (2005, 2010) argues that expected utility of income models are reference dependent like prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1992; Schmidt et al. 2008), and hence a major breakaway from the standard rational model. Empirical studies, however, typically do find risk aversion over modest stakes and it is commonly described using expected utility of income models. Some examples from the recent literature are Holt and Laury (2002), Harrison et al. (2007), and Andersen et al. (2008).

In this paper, I examine some consequences of defining utility on either income or initial wealth and income, rather than final wealth alone. The paper makes several contributions to the literature: a framework for describing expected utility models and preference reversals, a series of illustrations of preference reversals with expected utility of income models, and a theorem. I argue that whether or not income is an argument of utility has important consequences. If it is, then preferences are reference point dependent, since how outcomes are evaluated depends on what is counted as income and what is counted as wealth. The consequence of this reference dependence is that expected utility of income models can predict patterns of behavior usually thought of as anomalies for standard theory.

The framework to describe decision problems and expected utility models is developed in Section 1. It is based on the framework of states of the world, consequences and acts introduced by Savage (1954) and adapted by Sugden (2003). Preferences are defined over acts. Acts are also used to define what is counted as initial wealth and what is counted as income. The framework is also used to define different types of anomaly allowing the unified treatment of a range of anomalies.

[1] Other authors have also presented expected utility models that are explicitly not defined on wealth alone. An early example is Markowitz (1952) and a more recent one Sugden (2003).

[2] Another response to Rabin's paradox is the argument that other theories are vulnerable to analogous calibrations, so calibration may be a problem for all decision theories, not just expected utility of final wealth. Cox and Sadiraj (2006) present a concavity calibration proposition for small and large stake risk aversion that applies to expected utility of income models and some non-expected utility theories. Rubinstein (2006) considers time preferences and presents a calibration showing how constant discounting and seemingly plausible inter-day discounting predict implausible degrees of discounting over longer periods.

Different types of reference point effect can occur with expected utility of income models. In Section 2, a number of examples are presented, adapting several previously reported results. First, certain models predict framing induced preference reversals. This occurs when changing the reference point changes the preference order of two options. Changing the reference point can be compared to changing the perspective on a visual scene. A framing induced preference reversal is like changing the position from which two mountains are viewed changing which one appears highest.[3] Second, the models can predict a willingness to pay-willingness to accept gap for lotteries. That is, the maximum amount a person would be willing to pay to obtain a lottery is less than the minimum they would be willing to accept to give it up had they owned it. Third, the models can predict choice-value preference reversals. These occur when one of a pair of lotteries is chosen in a straight choice but the other is valued higher when certainty equivalents of the two lotteries are elicited.

One might wonder if these preference reversals only occur with certain utility functions or assumptions about how outcomes are coded in terms of initial wealth and income. The main contribution of this paper is a theorem, which is presented in Section 3. It is proved that for all utility functions defined on income or initial wealth and income, there will be framing induced preference reversals unless the preferences can be described by a utility function defined on final wealth alone. The intuition for the proof is that changing the reference point never changes preferences if and only if an expected utility of final wealth model can represent the preferences. The wider implications for decision theory are discussed in Section 4. One important implication is that since empirical studies typically do find risk aversion over modest stakes, the possibility of preference reversals occurring should be taken seriously when modeling individual behavior and strategic interactions between agents.

1 Framework

A framework based on the one developed by Savage (1954) and adapted by Sugden (2003) is used to describe expected utility models, decision problems, and anomalies. There is a set, S, of mutually exclusive **states of the world**. States represent different resolutions of uncertainty. The number of states is finite. The probability of state s obtaining is $p(s)$ where $p : S \to [0, 1]$. For all $s \in S$, $p(s) > 0$. An **act**, $\mathbf{a}$ is a function mapping from S to **consequences** C, $\mathbf{a} : S \mapsto C$. The set of all possible acts is denoted by $\mathbf{A}$. A special case is a **constant act**. It gives the same consequence in all states of the world, that is for all $s, s' \in S$, $a(s) = a(s')$, and represents a sure amount. Acts serve two purposes in the framework. First, acts are the objects of choice. Second, acts are used to define the initial wealth of a decision maker and so determine what part of the outcome is counted as income.

[3]This analogy was introduced by Tversky and Kahneman (1981).

In this paper consequences are restricted to levels of wealth. Let $C_w = \{w \in \mathbb{R} : w \geq 0\}$ be a set of non negative real numbers representing wealth levels.

Definition 1 A **utility of final wealth function** is $u_w : C_w \to \mathbb{R}$ where for all $w, w' \in C_w$, if $w > w'$ then $u_w(w) > u_w(w')$. The set of all such functions is denoted U_w.

The utility functions in U_w can be used to construct corresponding expected utility models.

$$EU_w(\mathbf{a}) = \sum_{s \in S} p(s) u_w(a(s))$$

Act **a** is preferred to act **b** if and only if $EU_w(\mathbf{a}) > EU_w(\mathbf{b})$. Preference is a binary relation between **a** and **b**.

To model utility of income, final wealth w is disaggregated into initial wealth r and income y. Let $D_y = \{y \in \mathbb{R} : r \geq 0 \wedge (r + y) \in C_w\}$.

Definition 2 A **utility of income function** is $u_y : D_y \to \mathbb{R}$ where for all $y, y' \in D_y$, if $y > y'$ then $u_y(y) > u_y(y')$. The set of all such functions is denoted U_y.

To model expected utility of initial wealth and income, let a set of ordered initial wealth-income pairs $D_{ry} = \{(r, y) \in \mathbb{R}^2 : r \geq 0 \wedge (r + y) \in C_w\}$.

Definition 3 A **utility of initial wealth and income function** is $u_{ry} : D_{ry} \to \mathbb{R}$ where for all (r, y), $(r', y') \in D_{ry}$, if $r = r'$ and $y > y'$ then $u_{ry}(r, y) > u_{ry}(r', y')$. The set of all such functions is denoted U_{ry}.

An act **a** gives final wealth in each state of the world. Income, y is the difference between initial and final wealth. Let initial wealth be defined by the act **r**. Income in state s is $a(s) - r(s)$.

When **r** is a constant act, expected utility of income for utility function u_y is defined as follows.

$$EU_y(\mathbf{r}, \mathbf{a}) = \sum_{s \in S} p(s) u_y(a(s) - r(s))$$

Expected utility of initial wealth and income for utility function u_{ry} is defined in a similar way.

$$EU_{ry}(\mathbf{r}, \mathbf{a}) = \sum_{s \in S} p(s) u_{ry}(r(s), a(s) - r(s))$$

For both EU_y and EU_{ry} models, when initial wealth is given by act **r**, act **a** is preferred to act **b** if and only if $EU(\mathbf{r}, \mathbf{a}) > EU(\mathbf{r}, \mathbf{b})$. When a set of options is compared, the same act **r** defining initial wealth is used for all the options. Preference is a triadic relation between **a**, **b**, and **r**.

The framework allows initial wealth to be uncertain. When **r** is not a constant act, there are a number of ways expected utility of income and expected utility of initial wealth and income can be modeled. A simple approach is a state-by-state

comparison of $\mathbf{r}$ and $\mathbf{a}$ with $EU_y(\mathbf{r}, \mathbf{a})$ and $EU_{ry}(\mathbf{r}, \mathbf{a})$ defined by the equations above.[4]

The framework and the differences between the three expected utility models can be illustrated using game of roulette. An American roulette wheel has 38 numbered pockets where the ball can land. When a player bets on a single number, they receive 36 times the stake if the ball lands on the number and zero otherwise. Suppose the player has a certain initial wealth of \$100 when entering the casino. This can be represented by the constant act $\mathbf{w_{100}}$. The set of states of the world contains one state for each pocket. The set of acts that are chosen between are the admissible bets. The act of betting one dollar on number 8, denoted $\mathbf{b_8}$, gives the consequence of \$100 − \$1 + \$36 = \$135 if the ball lands on number 8 and \$100 − \$1 = \$99 otherwise. The expected utility of placing the bet in each of the three models is given by the following equations.

$$\begin{aligned} EU_w(\mathbf{b_8}) &= \frac{1}{38}u_w(135) + \frac{37}{38}u_w(99) \\ EU_y(\mathbf{w_{100}}, \mathbf{b_8}) &= \frac{1}{38}u_y(135 - 100) + \frac{37}{38}u_y(99 - 100) \\ EU_{ry}(\mathbf{w_{100}}, \mathbf{b_8}) &= \frac{1}{38}u_{ry}(100, 135 - 100) + \frac{37}{38}u_{ry}(100, 99 - 100) \end{aligned}$$

Notice that although the utility functions u_y and u_{ry} take income as an argument, EU_y and EU_{ry} take acts as arguments and acts always map from S to C_w.

The example can be extended to illustrate using an uncertain act to define initial wealth. Suppose one dollar has been bet on 8 but the roulette wheel has not been spun yet and the player considers placing an additional dollar bet on 8, denoted by the act $\mathbf{b'_8}$. The act $\mathbf{b_8}$ could be taken as the initial wealth, in which case the expected utility of wealth and income using the state-by-state approach would be as follows.

$$EU_{ry}(\mathbf{b_8}, \mathbf{b'_8}) = \frac{1}{38}u_{ry}(135, 170 - 135) + \frac{37}{38}u_{ry}(99, 98 - 99)$$

The extended example can also illustrate the difficulties inherent in deciding which consequences are initial wealth. An alternative act (either constant or uncertain) could be taken as the initial wealth which could give rise to different decisions. For instance, if act $\mathbf{w_{100}}$ defined initial wealth, the expected utility from betting an extra dollar on number 8 would be $EU_{ry}(\mathbf{w_{100}}, \mathbf{b'_8})$ rather than $EU_{ry}(\mathbf{b_8}, \mathbf{b'_8})$.

Preference reversals resulting from changing the initial wealth can be defined as follows.

Definition 4 A **framing induced preference reversal** occurs if there exist $\mathbf{r}, \hat{\mathbf{r}}, \mathbf{a}, \mathbf{b} \in \mathbf{A}$ such that

$$EU(\mathbf{r}, \mathbf{a}) > EU(\mathbf{r}, \mathbf{b})$$

[4]In Section 2.2 the state-by-state approach is compared to an approach similar to that taken by Koszegi and Rabin (2007).

but

$$EU(\hat{\mathbf{r}}, \mathbf{a}) \leq EU(\hat{\mathbf{r}}, \mathbf{b}).$$

This definition is used in the following section and in the theorem reported in Section 3. Two additional types of preference reversal, reference-dependent-valuations and choice-valuation preference reversals, are also defined and discussed in the next section.

2 Anomalies under expected utility of income models

Expected utility of income models can predict preference reversals. This is illustrated with the following utility function, where y is income in dollars (i.e. losses or gains in wealth).

$$u_y(y) = \begin{cases} 0.9y + 0.1 & y < 1 \\ y^{0.9} & y \geq 1 \end{cases} \tag{1}$$

This utility function was proposed by Cox and Sadiraj to show that an expected utility of income model can explain risk aversion over modest stakes without implying absurd risk aversion over large stakes.[5] Notice that although the function is defined on losses and gains like prospect theory's value function, there is no kink at the origin.

2.1 Framing induced preference reversals

Expected utility of income models can predict framing induced preference reversals. That is, for some utility functions defined on income, changing how outcomes are coded in terms of initial wealth and income changes preferences. Table 1 shows two examples illustrating this. The first example is shown in the top section of the table. With initial wealth \$200, an expected utility of income maximizer whose preferences are described by Eq. 1 prefers A to B. With initial wealth \$400, they prefer D to C. Notice, however, that option A gives the same final wealth as C; B gives the same as D. Changing the initial wealth from \$200 to \$400 has reversed their preference over the two outcomes defined in terms of final wealth.[6]

[5]The potential for anomalies does not go unnoticed by Cox and Sadiraj. For instance in their footnote 8 they write that the expected utility of initial wealth and income model they introduce "does not rule out certain types of anomalies (see Rubinstein 2006 for an illustration). Detailed analysis of possible "money pump" preference cycles and other violations of full rationality are beyond the scope of the present paper, which is concerned with the implications of concavity calibration for decision theories."

[6]This example is a slight modification of one Kahneman and Tversky (1982) use. It shows how the same decision problem can be framed in different ways and how different frames can lead people to choose different options.

Table 1 Framing induced preference reversals

Initial Wealth	Option	Income	Expected Utility	Final Wealth
Utility function: Equation 1				
$200	A	$49	33.2	$249
$200	B	($200, .25; $0, .75)	29.5	($400, .25; $200, .75)
$400	C	−$151	−135.8	$249
$400	D	($0, .25; −$200, .75)	−134.9	($400, .25; $200, .75)
Utility function: Equation 2				
$50	E	$180	25.0	$230
$50	F	($350, .25; $150, .75)	25.1	($400, .25; $200, .75)
$150	G	$80	17.6	$230
$150	H	($250, .25; $50, .75)	17.2	($400, .25; $200, .75)

In the "Income" and "Final Wealth" columns, lotteries are described as a series of prize-probability pairs. For example, ($200, .25; $0, .75) means $200 with probability .25 and zero with probability .75. For sure amounts, probability values are omitted. In the top section of the table, the expected utility values are calculated using Eq. 1 as the utility function; in the bottom section, they are calculated using Eq. 2

Framing induced preference reversals can occur even when all the outcomes being evaluated are above the initial wealth. This is illustrated in a second example that uses the following utility function.

$$u_y(y) = \frac{1 - exp(-\alpha y^{1-\beta})}{\alpha} \tag{2}$$

This function was used by Holt and Laury (2002) as part of a stochastic choice model. Using data from a lottery-choice experiment they estimated the two parameters as $\alpha = 0.029$ and $\beta = 0.269$. The example, using the utility function and estimated parameters, is shown in the bottom section of the table. With initial wealth $50, the risky option F is preferred to the safe option E. With initial wealth $150, the safe option G is preferred to the risky option H.

2.2 Reference-dependent valuations

The framework can be used to describe reference dependent valuations for lotteries.

Definition 5 Valuations are **reference dependent** if there exist $\mathbf{r}, \hat{\mathbf{r}}, \mathbf{a}, \mathbf{v}, \underline{\mathbf{v}} \in \mathbf{A}$ such that

$$EU(\mathbf{r}, \mathbf{a}) = EU(\mathbf{r}, \mathbf{v})$$

and

$$EU(\hat{\mathbf{r}}, \mathbf{a}) = EU(\hat{\mathbf{r}}, \underline{\mathbf{v}})$$

but for all $s \in S$, $v(s) \geq \underline{v}(s)$ and for some $s \in S$, $v(s) > \underline{v}(s)$.

Reference dependent valuations are usually thought of as an anomaly for standard theory. For instance, List (2003) writes:

> the basic independence assumption, which is used in most theoretical and applied economic models to assess the operation of markets, has been directly refuted in several experimental settings [Knetsch 1989; Kahneman et al. 1990; Bateman et al. 1997]. These experimental findings have been robust across unfamiliar goods, such as irradiated sandwiches, and common goods, such as chocolate bars, with most authors noting behavior consistent with an endowment effect. Such findings have induced even the most ardent supporters of neoclassical theory to doubt the validity of certain neoclassical modeling assumptions.

Although the endowment effect is commonly thought of as a phenomenon in riskless choice, it can also occur in risky settings. There are several ways specific measures of value can be defined. I focus on four measures Bateman et al. (1997) define, and apply them to a choice between a binary lottery and a sure amount. Some pairs of measures can be used to identify reference dependence as defined in Definition 5. For simplicity, it is assumed there is no background risk.[7] Let $\underline{\mathbf{w}}$ be a constant act that gives x in every state. Let $\mathbf{b}$ be an act representing playing the lottery. It gives $x + prize$ where $prize$ is the lottery payout in the state. For each of the cases below, the valuation measure is the amount of money that makes a person indifferent between the two options. For two of the measures, initial wealth is the lottery $\mathbf{b}$ and for two it is the constant act $\underline{\mathbf{w}}$. In some cases, an amount is added or subtracted from one of the acts. For example, $\mathbf{b} - WTP$ means the consequence in state s is $b(s) - WTP$.

Willingness to accept (WTA): A person is endowed with $\mathbf{b}$ and is indifferent between (a) selling $\mathbf{b}$ for WTA or (b) keeping $\mathbf{b}$.

$$EU(\mathbf{b}, \mathbf{b}) = EU(\mathbf{b}, \underline{\mathbf{w}} + WTA)$$

Willingness to pay (WTP): A person is indifferent between (a) buying $\mathbf{b}$ for WTP or (b) not trading.

$$EU(\underline{\mathbf{w}}, \underline{\mathbf{w}}) = EU(\underline{\mathbf{w}}, \mathbf{b} - WTP)$$

Equivalent gain (EG): A person is indifferent between (a) gaining $\mathbf{b}$ or (b) gaining EG.

$$EU(\underline{\mathbf{w}}, \mathbf{b}) = EU(\underline{\mathbf{w}}, \underline{\mathbf{w}} + EG)$$

Equivalent loss (EL): A person is endowed with $\mathbf{b}$ and is indifferent between (a) giving up $\mathbf{b}$ or (b) giving up EL.

$$EU(\mathbf{b}, \underline{\mathbf{w}}) = EU(\mathbf{b}, \mathbf{b} - EL)$$

[7] It would be a simple extension to include an act representing background risk.

Bateman et al. (1997) argue that the standard theory while not predicting equality between WTA and WTP does predict $EG = WTA$ and $EL = WTP$. Under reference dependent preferences, they write:

> If losses loom larger than gains, then, in the absence of income effects, we should expect WTA to be greater than WTP. EG, which expresses an equivalence between gains on the two dimensions, and EL which expresses an equivalence between losses, should be expected to take values intermediate between WTP and WTA.

The same result holds when Definition 5 is used to identify reference dependence. Valuations are reference dependent if $EG \neq WTP$ or $EL \neq WTA$. Table 2 shows WTA, WTP, EG and EL valuations for a binary lottery that pays out \$1000 with probability 0.3 and zero otherwise. There are several ways income could be measured when an agent holds a risky position. The framework described in Section 1 is used to describe two alternative approaches. Figures calculated using these two approaches are shown in the table.

First a state-by-state comparison as used by Sugden (2003), which was introduced earlier. For each state of the world, the consequence of the act **a** is compared to the consequence of the reference act **r** in that state.

$$EU(\mathbf{r}, \mathbf{a}) = \sum_{s \in S} p(s) u_y(a(s) - r(s)) \tag{3}$$

Second, an approach similar to that taken by Koszegi and Rabin (2007).[8] For each state of the world, the consequence of the act **a** is compared to the consequence of the reference act **r** in *every* state.

$$EU(\mathbf{r}, \mathbf{a}) = \sum_{s \in S} \sum_{t \in S} p(s) p(t) u_y(a(s) - r(t)) \tag{4}$$

Notice that when **r** is a constant act, the two approaches are equivalent, and so both are consistent with the definition of expected utility of initial wealth and income in Section 1.

The four value measures are calculated using the utility of income function u_y defined in Eq. 1. Notice that (a) WTA exceeds WTP and (b) both $EG \neq WTA$ and $EL \neq WTP$. This holds for both specifications of the reference point. The initial wealth is a sure amount for EG and WTP so (when there is no background risk) the state-by-state and the Koszegi-Rabin approaches give the same result. When the initial wealth is the lottery, the two approaches give different results. For the state-by-state approach, WTA exceeds expected value (the lottery's expected value is 300). For the Koszegi-Rabin approach, all measures are less than expected value.

Consider the consequences of the differences in the valuations for a variant of Knetsch's 1989 chocolate and mugs experiment. Suppose a person with preferences

[8] Other aspects of Koszegi and Rabin's model (such as separating standard consumption utility and loss gain utility and making the reference point a person's recent rational expectations about outcomes) are not used in this paper.

Table 2 Reference dependent valuations

Measure	State-by-state	Koszegi-Rabin
Willingness to accept (WTA)	\$413.4	\$278.6
Equivalent loss (EL)	\$300.0	\$230.8
Equivalent gain (EG)	\$262.6	\$262.6
Willingness to pay (WTP)	\$196.1	\$196.1

The table shows valuations for the same lottery calculated using the four measures and the two approaches described in this section. The valuations are for a lottery that pays out \$1000 with probability 0.3 and zero otherwise.

as in Table 2 is given the lottery and then asked if they want to swap it for \$250. For the lottery $WTA > 250$ so the trade will be refused. Now suppose instead the person had been endowed with 250 and asked if they want to swap it for the lottery. For the lottery $WTP < 250$ so the trade will be refused. There would be an endowment effect.

2.3 Choice-valuation preference reversals

Standard economic theory implies that the preference ordering of a pair of alternatives should not depend on the process used to elicit it. There is a large body of evidence, however, suggesting that for certain classes of lottery, the process used to elicit preferences systematically alters the rank ordering. In many experimental studies when people evaluate a "P-bet" (a lottery that pays a modest amount with a high probability) and a "\$-bet" (a lottery that pays out a high amount with a small probability), people tend to choose the P-bet but report a higher WTA value for the \$-bet (see Cubitt et al. (2004) for a review of studies). This is problematic for standard theory. For instance, Grether and Plott (1979) write:

> Taken at face value the data are simply inconsistent with preference theory and have broad implications about research priorities within economics. The inconsistency is deeper than the mere lack of transitivity or even stochastic transitivity. It suggests that no optimization principles of any sort lie behind even the simplest of human choices...

Choice-valuation preference reversals such as commonly observed with the P-bet and \$-bet can be described using the framework from Section 1.

Definition 6 There is a **choice-valuation preference reversal** if there exist acts $\mathbf{r}, \hat{\mathbf{r}}, \tilde{\mathbf{r}}, \mathbf{a}, \mathbf{b}, \mathbf{v}, \underline{\mathbf{v}} \in \mathbf{A}$ such that

$$EU(\mathbf{r}, \mathbf{a}) > EU(\mathbf{r}, \mathbf{b})$$

but

$$EU(\hat{\mathbf{r}}, \mathbf{a}) = EU(\hat{\mathbf{r}}, \underline{\mathbf{v}})$$

Table 3 Choice-valuation preference reversals

Measure	State-by-state		Koszegi-Rabin	
	P-bet	$-bet	P-bet	$-bet
WTA	$305.1	$478.2	$282.7	$284.6
$EU(0, \mathbf{b})$	157.8	153.6	157.8	153.6

The P-bet pays out 350 with probability 0.81 and zero otherwise. The $-bet pays out 1700 with probability 0.19 and zero otherwise. WTA is willingness to accept. $EU(0, \mathbf{b})$ is the expected utility of the respective lottery when initial wealth is zero. The figures are calculated using the utility function defined in Eq. 1

and

$$EU(\tilde{\mathbf{r}}, \mathbf{b}) = EU(\tilde{\mathbf{r}}, \mathbf{v})$$

where for all $s \in S$, $v(s) \geq \underline{v}(s)$ and for some $s \in S$, $v(s) > \underline{v}(s)$.

This definition can be applied to the P-bet/$-bet problem as follows. Act **a** represents the P-bet and act **b** the $-bet. Act $\underline{\mathbf{v}}$ is the WTA for the P-bet and act **v** is the WTA for the $-bet. Finally act **r** is the constant act that gives zero in every state, act $\hat{\mathbf{r}}$ is the P-bet and act $\tilde{\mathbf{r}}$ is the $-bet.

Table 3 shows WTA valuations and expected utility figures for a P-bet and a $-bet. The P-bet pays 350 with probability 0.81; the $-bet pays out 1700 with probability, 0.19. The WTA valuation for the $-bet is higher, suggesting the $-bet is preferred. The expected utility figure, however, is higher for the P-bet meaning that in a straight choice, the P-bet is preferred. The preference inferred from the WTA valuations is reversed.[9]

3 Generalization

The previous sections considered a utility function defined on income alone. This section considers models where utility can depend on both initial wealth and income.

Theorem 1 *For all expected utility of initial wealth and income models with utility function $u_{ry} \in U_{ry}$ either (a) there are preference reversals or (b) preferences can be represented using an expected utility of final wealth model with utility function $u_w \in U_w$.*

Proof Let **A** be the set of all possible acts. Take any utility function $u_{ry} \in U_{ry}$. Let EU_{ry} be an expected utility of initial wealth and income model using utility function u_{ry}. Take any pair of acts $\mathbf{a}, \mathbf{b} \in \mathbf{A}$, and any initial wealth $\mathbf{r} \in \mathbf{A}$. From initial

[9]That choice-value preference reversals can occur in an expected utility model is not a new result. For instance, Sugden (2003) shows how similar results occur in an expected utility model where utility is defined on satisfaction and changes in satisfaction.

wealth $\mathbf{r}$, either (1) $EU_{ry}(\mathbf{r}, \mathbf{a}) > EU_{ry}(\mathbf{r}, \mathbf{b})$, (2) $EU_{ry}(\mathbf{r}, \mathbf{a}) < EU_{ry}(\mathbf{r}, \mathbf{b})$, or (3) $EU_{ry}(\mathbf{r}, \mathbf{a}) = EU_{ry}(\mathbf{r}, \mathbf{b})$. Consider cases (1) and (2).

Case 1. If there is at least one framing induced preference reversal, then there exists some $\hat{\mathbf{r}} \in \mathbf{A}$ such that $EU_{ry}(\hat{\mathbf{r}}, \mathbf{a}) \leq EU_{ry}(\hat{\mathbf{r}}, \mathbf{b})$. Conversely, if there are no such preference reversals, then for all $\hat{\mathbf{r}} \in \mathbf{A}$, $EU_{ry}(\hat{\mathbf{r}}, \mathbf{a}) > EU_{ry}(\hat{\mathbf{r}}, \mathbf{b})$.

Case 2. If there is at least one framing induced preference reversal, then there exists some $\hat{\mathbf{r}} \in \mathbf{A}$ such that $EU_{ry}(\hat{\mathbf{r}}, \mathbf{a}) \geq EU_{ry}(\hat{\mathbf{r}}, \mathbf{b})$. Conversely, if there are no such preference reversals, then for all $\hat{\mathbf{r}} \in \mathbf{A}$, $EU_{ry}(\hat{\mathbf{r}}, \mathbf{a}) < EU_{ry}(\hat{\mathbf{r}}, \mathbf{b})$.

Now, assume there are no framing induced preference reversals with EU_{ry} for all $\mathbf{r}, \hat{\mathbf{r}}, \mathbf{a}, \mathbf{b} \in \mathbf{A}$. Let $\mathbf{r_0} \in \mathbf{A}$ be the constant act the gives zero in all states. Since there are no preference reversals, it follows that (1) $EU_{ry}(\mathbf{r}, \mathbf{a}) < EU_{ry}(\mathbf{r}, \mathbf{b})$ if and only if $EU_{ry}(\mathbf{r_0}, \mathbf{a}) < EU_{ry}(\mathbf{r_0}, \mathbf{b})$ and (2) $EU_{ry}(\mathbf{r}, \mathbf{a}) > EU_{ry}(\mathbf{r}, \mathbf{b})$ if and only if $EU_{ry}(\mathbf{r_0}, \mathbf{a}) > EU_{ry}(\mathbf{r_0}, \mathbf{b})$. From this it follows that if $EU_{ry}(\mathbf{r}, \mathbf{a}) = EU_{ry}(\mathbf{r}, \mathbf{b})$, then neither (1) $EU_{ry}(\mathbf{r_0}, \mathbf{a}) < EU_{ry}(\mathbf{r_0}, \mathbf{b})$ nor (2) $EU_{ry}(\mathbf{r_0}, \mathbf{a}) > EU_{ry}(\mathbf{r_0}, \mathbf{b})$. Hence, $EU_{ry}(\mathbf{r}, \mathbf{a}) = EU_{ry}(\mathbf{r}, \mathbf{b})$ if and only if $EU_{ry}(\mathbf{r_0}, \mathbf{a}) = EU_{ry}(\mathbf{r_0}, \mathbf{b})$.

Recall that

$$EU_{ry}(\mathbf{r_0}, \mathbf{a}) = \sum_{s \in S} p(s) u_{ry}(0, a(s)).$$

where $u_{ry} : D_{ry} \to \mathbb{R}$. Since for all $(0, x) \in D_{ry}$ there exists $x \in C$, it follows that for all $u_{ry} \in U_{ry}$, there exists $u_w \in U_w$ such that for all $x \in C$, $u_{ry}(0, x) = u_w(x)$. Hence, there exists $u_w \in U_w$ such that

$$\sum_{s \in S} p(s) u_{ry}(a(s), 0) = \sum_{s \in S} p(s) u_w(a(s)).$$

Hence, either there are preference reversals, or preferences can be captured by a utility function defined on final wealth alone. □

It is noteworthy that it has not been proved that all expected utility of income models predict preference reversals.

Figure 1 is a Venn diagram showing the relations between expected utility models. Each point on the diagram is a preference relation over lotteries. There are three sets labeled R_w, R_y, and R_{ry}. The sets contain preference relations described by expected utility models defined on (i) final wealth, (ii) income, and (iii) initial wealth and income. The shaded areas represent empty zones. Let w denote final wealth, r denote initial wealth, and y denote income. The points labeled a, b and c represent different cases. It is easy to see that preferences represented by expected utility of wealth models (point a) and by expected utility of income models (point b) can also be represented by an initial wealth and income model by simply using a utility function $u_{ry}(r, y) \equiv u_w(r + y)$ or $u_{ry}(r, y) \equiv u_y(y)$ respectively. More interesting are the preferences represented by point c, which can be described by all three classes of model. Such preferences can be described by an exponential utility function such as $u(x) = -e^{-x}$, which, as Pratt (1964) notes, imply the same risk preferences at all levels of x. Such preferences exhibit constant absolute risk aversion (CARA). Utility functions can be defined on income or final wealth without preference reversals. Although CARA preferences have attractive features, many empirical

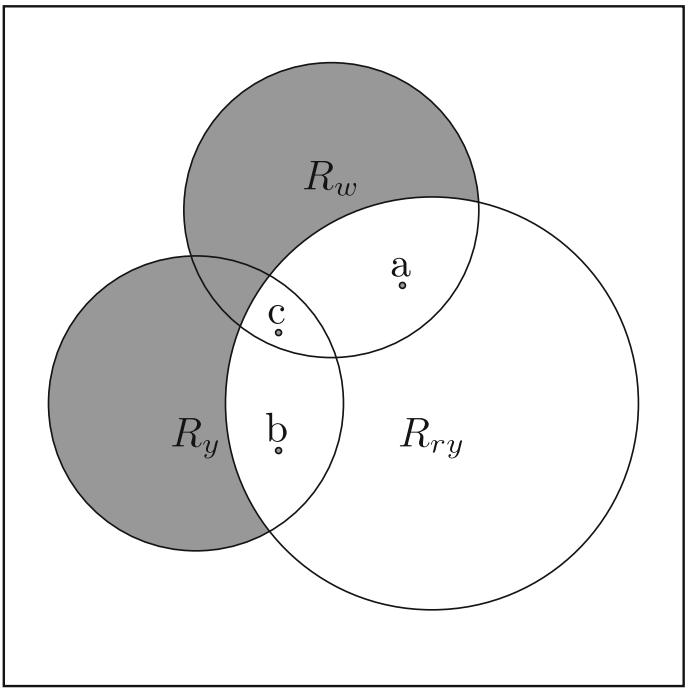

Fig. 1 Relations between expected utility models. R_w is the set of preference relations described by expected utility of wealth models. R_y is the set of preference relations described by expected utility of income models. R_{ry} is the set of preference relations described by expected utility of initial wealth and income models. Shaded areas indicate empty zones

studies use a utility function exhibiting constant relative risk aversion.[10] Such functions allow a person's preferences over risks to change with the assets they hold, so as a consequence must be defined on final wealth if preference reversals are to be avoided.

4 Implications for decision theory

This paper has shown that if income and wealth are separate arguments of utility, the resulting model can predict anomalies that are inconsistent with the standard rational model. But if income and wealth do not enter separately, then Rabin's paradox applies and risk aversion over modest stakes cannot be accommodated. Expected utility of initial wealth and income models cannot accommodate risk aversion over modest stakes without also accommodating preference reversals.

Empirical studies typically find risk aversion over modest stakes, which suggests we should take seriously the possibility of preference reversals occurring. There are several important open questions. If preferences are reference dependent, how is the

[10] See Wakker (2008) for a discussion of the characteristics of such functions.

reference point determined and what dynamic inconsistencies does this cause? Do people anticipate dynamic inconsistencies in their own behavior (for example as in the dual self model of Fudenberg and Levine (2006)), and if so, how do they react? Do people anticipate dynamic inconsistency in the behavior of others and what are the consequences for strategic interaction?

Acknowledgments I am grateful to Thomas Epper, Jacob Goeree, Konrad Mierendorff, Chris Starmer, Jingjing Zhang, and participants at the FUR XIV International Conference at Newcastle University, as well as the editor and one anonymous referee, for comments. I would like to thank the Swiss National Science Foundation (grant SNSF 138162) and the European Research Council (grant ESEI-249433) for financial support.

References

Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E.E. (2008). Eliciting risk and time preferences. *Econometrica*, *76*, 583–618.

Bateman, I., Munro, A., Rhodes, B., Starmer, C., Sugden, R. (1997). A test of the theory of reference-dependent preferences. *Quarterly Journal of Economics*, *112*, 479–505.

Cox, J.C., & Sadiraj, V. (2006). Small-and large-stakes risk aversion: Implications of concavity calibration for decision theory. *Games and Economic Behavior*, *56*, 45–60.

Cubitt, R.P., Munro, A., Starmer, C. (2004). Testing explanations of preference reversal. *The Economic Journal*, *114*, 709–726.

Fudenberg, D., & Levine, D.K. (2006). A dual-self model of impulse control. *American Economic Review*, *96*, 1449–1476.

Grether, D.M., & Plott, C.R. (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, *69*, 623–638.

Harrison, G.W., List, J.A., Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, *75*, 433–458.

Holt, C.A., & Laury, S.K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*, 1644–1655.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, *246*, 160–173.

Kahneman, D., Knetsch, J.L., Thaler, R.H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, *98*(6), 1325–1348.

Knetsch, J.L. (1989). The endowment effect and evidence of nonreversible indifference curves. *American Economic Review*, *79*, 1277–1284.

Koszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, *97*, 1047–1073.

List, J.A. (2003). Does market experience eliminate market anomalies? *The Quarterly Journal of Economics*, *118*, 41–71.

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy*, *60*, 151.

Pratt, J.W. (1964). Risk aversion in the small and in the large. *Econometrica*, *32*, 122–136.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, *68*, 1281–1281.

Rabin, M., & Thaler, R.H. (2001). Anomalies: Risk aversion. *The Journal of Economic Perspectives*, *15*, 219–232.

Rubinstein, A. (2006). Dilemmas of an economic theorist. *Econometrica*, *74*, 865–883.

Savage, L.J. (1954). *The foundation of statistics*. New York: Wiley.

Schmidt, U., Starmer, C., Sugden, R. (2008). Third-generation prospect theory. *Journal of Risk and Uncertainty*, *36*, 203–223.

Sugden, R. (2003). Reference-dependent subjective expected utility. *Journal of Economic Theory*, *111*, 172–191.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.
von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton: Princeton University Press.
Wakker, P.P. (2005). Formalizing reference dependence and initial wealth in Rabin's calibration theorem. http://people.few.eur.nl/wakker/pdf/calibcsocty05.pdf.
Wakker, P.P. (2008). Explaining the characteristics of the power (CRRA) utility family. *Health Economics*, *17*, 1329–1344.
Wakker, P.P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge: Cambridge University Press.