Paradoxes and Mechanisms for Choice under Risk

By James C. Cox, Vjollca Sadiraj, and Ulrich Schmidt[1]

Abstract: Experiments on choice under risk typically involve multiple decisions by individual subjects. The choice of mechanism for selecting decision(s) for payoff is an essential design feature unless subjects isolate each one of the multiple decisions. We discuss theoretical properties of commonly used mechanisms and new mechanisms introduced herein. We report experiments that generate data that show systematic differences across mechanisms in subjects' revealed risk preferences. We illustrate the importance of these mechanism effects by identifying their implications for tests of classic properties of theories of decision under risk. We also identify behavioral properties of mechanisms that introduce bias in elicited risk preferences from cross-task contamination.

Keywords: experiments, risk, payoff mechanisms, paradoxes, cross-task contamination
JEL classifications: C91, D81

1. Introduction

Most experiments on choice under risk involve multiple decisions by individual subjects. This necessitates choice of mechanism for determining incentive payments to the subjects. Mechanisms used in papers published by top five general readership journals and a prominent field journal vary quite widely from "paying all decisions sequentially" to "paying all decisions at the end" to "randomly paying one decision for each subject" to "randomly paying a few decisions for each subject" to "randomly paying some of the subjects" to "randomly paying one of the subjects" to "fixed payment" to unidentified mechanisms.[2] This suggests questions about whether different payoff mechanisms can elicit different data in otherwise identical experimental treatments and, if so, whether these mechanism effects have significant implications for conclusions drawn from data. We report an experiment with several payoff mechanisms that directly addresses these questions. Data from our experiment show that subjects' revealed risk preferences differ across mechanisms. We illustrate the importance of

[1] Financial support was provided by the National Science Foundation (grant number SES-0849590) and the Fritz Thyssen Stiftung. Valuable comments and suggestions were provided by the referees and editor and Glenn W. Harrison.

[2] Table 1 in Azrieli, et al. (2012) reports a survey of some of the payoff mechanisms used in papers published in 2011 in *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Review of Economic Studies*, *Quarterly Journal of Economics*, and *Experimental Economics*.

these payoff mechanism effects by using data from alternative mechanisms to test for consistency with classic properties of theories of decision under risk.

We provide an explanation of theoretical incentive compatibility or incompatibility of alternative mechanisms for decision theories with functionals that are linear in probabilities or linear in payoffs or linear in neither or linear in both. Data from our experiments are used to identify mechanism biases in risk preference elicitation such as choice-order effects and other types of cross-task contamination in which a subject's answer in one decision task may be affected by the content of some other decision tasks or the choices made in some other tasks.

2. Classic Properties of Theories of Decision under Risk

Allais (1953) first raised an objection to the independence axiom of expected utility theory by constructing thought experiments that seem to imply paradoxical outcomes. Subsequent behavioral experiments focused on two patterns that contradict the independence axiom, the common ratio effect (CRE) and common consequence effect (CCE). As we shall explain, some of the lottery pairs used in our experiment were selected because they make it possible to observe CRE and CCE if they characterize experimental subjects' revealed risk preferences.

Yaari (1987) introduced the dual independence axiom and constructed an alternative theory with functional that is nonlinear in probabilities (unless the agent is risk neutral) and linear in payoffs (for all risk attitudes). The dual common ratio effect (DCRE) and dual common consequence effect (DCCE) are the dual analogs of CRE and CCE. Some of the lottery pairs used in our experiment were selected because they make it possible to observe DCRE and DCCE if they characterize experimental subjects' revealed risk preferences.

A test for CRE uses two lottery pairs where the lotteries in one pair (Pair 3 in Table 1) are constructed from the lotteries in the other pair (Pair 2 in Table 1) by multiplying all probabilities by a common factor (1/4 in our study) and assigning the remaining probability to a common outcome ($0 in our study). It follows from linearity in probabilities of the expected utility functional that an expected utility agent would choose either the safer lotteries in both pairs or the riskier ones.[3] Any mixed choices of the riskier lottery in Pair 2 or Pair 3 and the safer lottery in the other pair reveals CRE.

A test for CCE also uses two lottery pairs. Here, the lotteries in the second pair (Pair 4 in Table 1) are constructed from the lotteries in the first pair (Pair 3 in Table 1) by shifting

[3] Within any pair of lotteries, we call the safer (resp. riskier) lottery the one with the smaller (resp. larger) variance.

probability mass (75% in our study) from one common outcome ($0 in our study) to a different common outcome ($12 in our study). It is easy to verify (with the functional) that expected utility theory requires that either the safer lotteries be chosen in both pairs or the riskier ones be chosen in both pairs. Any mixed choices, riskier lottery in Pair 3 or Pair 4 and the safer lottery in the other pair, reveals CCE.

Table 1. Lottery Pairs Used in the Experiment

Pair	Safer		Riskier		
1	Balls 1-15	Balls 16-20	Balls 1-16	Balls 17-20	
	$0	$3	$0	$5	
2	Balls 1-20		Balls 1-4	Balls 5-20	
	$6		$0	$10	
3	Balls 1-15	Balls 16-20	Balls 1-16	Balls 17-20	
	$0	$6	$0	$10	
4	Balls 1-5	Balls 6-20	Ball 1	Balls 2-5	Balls 6-20
	$6	$12	$0	$10	$12
5	Balls 1-20		Balls 1-4	Balls 5-20	
	$18		$12	$22	

The null hypotheses that follow from the independence axiom of expected utility theory are that the proportion of choices of the safer option in Pair 3 should be the same as the proportions of choices of the safer options in Pairs 2 and 4:

Hypothesis 1: The proportions of choices of the safer option are the same for Pair 2 and Pair 3 (absence of CRE).

Hypothesis 2: The proportions of choices of the safer option are the same for Pair 3 and Pair 4 (absence of CCE).

One-sided alternatives to the above hypotheses are provided by fanning-out (Machina, 1982) and fanning-in (Neilson, 1992).[4] Subjects' revealed risk preferences under each mechanism can be used to test these hypotheses.

DCRE and DCCE play the same role for dual theory of expected utility (Yaari, 1987) as CRE and CCE do for expected utility theory. Because the dual theory functional is linear in payoffs, it exhibits constant absolute and constant relative risk aversion. Consequently, neither multiplying all outcomes in a lottery pair by a constant (DCRE: see Pairs 1 and 3 in Table 1, where the constant equals 2) nor adding a constant to all outcomes in a lottery pair (DCCE: see Pairs 2 and 5 where the constant equals $12) affects choices. Yaari (1987) stated that the dual paradoxes could be used to refute his theory analogously to the way in which CRE and CCE had been used to refute expected utility theory. As far as we know, however, the dual paradoxes have never been investigated in a systematic empirical test with a *theoretically incentive compatible* mechanism.

The null hypotheses that follow from the dual independence axiom (which implies linearity in payoffs) are that the proportion of choices of the safer option should be: (a) the same in Pairs 1 and 3; and (b) the same in Pairs 2 and 5. The null hypothesis of choices in Pairs 1 and 3 coming from the same distribution also follows from a power function for payoffs, with or without linearity in probabilities. On the other hand, the null hypothesis of choices in Pairs 2 and 5 revealing the same distribution is consistent with an exponential function for payoffs. Data from each mechanism can be used to conduct tests of the following hypotheses:

Hypothesis 3: The proportions of choices of the safer option are the same for Pair 1 and Pair 3 (absence of DCRE).

Hypothesis 4: The proportions of choices of the safer option are the same for Pair 2 and Pair 5 (absence of DCCE).

One-sided alternatives to Hypothesis 3 are given by decreasing relative risk aversion (DRRA) or increasing relative risk aversion (IRRA). One-sided alternatives to Hypothesis 4 are provided by decreasing absolute risk aversion (DARA) or increasing absolute risk aversion (IARA).

---

[4] In the data analysis section, whenever we have a one-sided alternative hypothesis we will use one-sided p-values to draw conclusions about the rejection of the null hypothesis.

3. Theoretical Properties of Incentive Mechanisms

We consider several payoff mechanisms commonly used for multiple decision experiments and new mechanisms introduced herein. We also use another "mechanism" in which each subject makes only one decision.

The payoff mechanism that appears to be most commonly used in experiments on individual choice in strategic settings (such as markets and public goods) is the one in which each decision is paid sequentially before a subsequent decision is made; we label this mechanism "pay all sequentially" (PAS).[5] Another way in which all decisions are paid is to pay all decisions at the end of the experiment with independent draws of random variables; we label this mechanism "pay all independently" (PAI). Another mechanism commonly used in experiments on decision under risk is to randomly select one decision for payoff at the end of the experiment. There are two ways in which this payoff mechanism is commonly used which differ in whether a subject is shown all lotteries before making any choices. In one version of the mechanism (e.g., Holt and Laury 2002, Starmer and Sugden 1991) a subject is shown all lotteries in advance before any choices are made; we label this version of the mechanism "pay one randomly with prior information" (PORpi). In an alternative version of this mechanism (e.g., Hey and Orme 1994, Hey and Lee 2005a, 2005b) a subject is shown each lottery pair for the first time just before a choice is made; we call this version of the mechanism "pay one randomly with no prior information" (PORnp). To our best knowledge, a new mechanism is to pay all decisions at the end of the experiment with one realization of the state of the world that determines payoffs; the theoretical properties of this mechanism are explained below (for comonotonic lotteries). There are two versions of this mechanism that differ in scale of payoffs. In one version, full payoff for all chosen lotteries is made according to one random draw at the end of the experiment; we label this mechanism "pay all correlated" (PAC). With N decisions, the scale of the payoffs with PAC, which is the same as with PAS and PAI, is N times the expected payoff with either version of POR. The alternative version, called PAC/N, pays 1/N of the payoffs for all chosen lotteries; this version of the mechanism has the same scale of payoffs as (both versions of) POR.

In a review of the experimental evidence on violations of expected utility, Cubitt, Starmer, and Sugden (2001) advocate use of between-subjects designs, in which each subject makes one choice, rather than within-subjects designs with multiple decisions. We implement this approach and compare the resulting data to the data elicited by several multiple decision

[5] As shown in Table 1 of Azrieli, et al. (2012), this mechanism was used in 27 out of the 42 papers in which the chosen mechanism was reported for 2011 publications in the six journals listed in footnote 2, above.

protocols using the above payoff mechanisms. We subsequently refer to the single decision per subject protocol as the “one task” (OT) mechanism.

3.1 Incentive Compatibility

A payoff mechanism is incentive compatible if it provides incentives for truthful revelation of preferences. We consider two definitions, “strong incentive compatibility” and “weak incentive compatibility”, which differ in generality of the assumption one makes about interaction between events within and outside an experiment.

In the context of an experiment on pairwise choice, by *strong* incentive compatibility we mean the following. Suppose that the researcher is interested in eliciting an individual’s preference over some Option a and Option b in an experiment. The individual’s preference for Option a or Option b within the experiment may depend on the prizes and probability distribution F of states of the world external to the experiment. Let this preference ordering be denoted by $\succeq_F$ and assume that F is independent of what happens in the experiment (because the experimenter has control internal to the lab but no control external to the lab). The purpose of an experiment is to learn whether $a \succeq_F b$ or $b \succeq_F a$ by observing incentivized choice(s) between Option a and Option b. Incentivizing choices involves use of a payoff mechanism that may create incentives for “untruthful” revelation of the preference $\succeq_F$ over Option a and Option b. Let $\succeq_F^M$ denote the individual’s preferences when choices are implemented with mechanism M. Now consider the choice between Option a and Option b in the context of additional choices (in the experiment) between some Option A_i and Option B_i, for i = 1, 2, …, n. We say that payoff mechanism M is strongly incentive compatible when $a \succeq_F^M b$ if and only if $a \succeq_F b$ for all possible specifications of the n alternative pairs of options. This concept of strong incentive compatibility is used for some of our positive results and all of our counterexamples to incentive compatibility of several payoff mechanisms in the following subsections.

We also use a definition of *weak* incentive compatibility. Again suppose the researcher is interested in eliciting an individual’s preference for Option a or Option b. The individual’s preference for Option a or Option b within the experiment may depend on the amount of his (fixed, certain) wealth w^o outside the experiment. Let this preference ordering be denoted by $a \succeq_{w^o} b$. Let $\succeq_{w^o}^M$ denote the individual’s preferences when choices are implemented with mechanism M. We say that payoff mechanism M is weakly incentive compatible when

$a \succeq_{w^o}^{M} b$ if and only if $a \succeq_{w^o} b$ for all possible specifications of the n alternative pairs of options. This concept of weak incentive compatibility is used for some of our positive results about incentive compatibility of payoff mechanisms. Clearly, if a payoff mechanism is strongly incentive compatible then it is also weakly incentive compatible.

In this paper lotteries will often be denoted by $(X_1,p_1;...;X_n,p_n)$, indicating that outcome X_s is obtained with probability p_s, for $s=1,...,n$. Outcome X_s can be a monetary amount or a lottery. In an experiment that includes n questions in which the subject has to choose between Options A_i and B_i, for $i=1,...,n$, the choice of the subject in question i will be denoted by C_i.

3.2 *The Pay One Randomly (PORnp and PORpi) Mechanisms*

With these mechanisms each decision usually has a $1/n$ chance of being played out for real. Consider a subject who conforms to the reduction of compound lotteries axiom and has made all her choices except the choice in task i. As discussed by Holt (1986), her choice between A_i and B_i determines whether she will receive compound lottery (A_i,1/n;C,1-1/n) or (B_i,1/n;C,1-1/n), where C = (C_1, 1/(n-1); …; C_{i-1}, 1/(n-1); C_{i+1}, 1/(n-1); …; C_n, 1/(n-1)) is the lottery for which the subject receives all her previous choices with equal probability 1/(n-1). Consequently, a subject whose preferences satisfy the reduction and independence axioms has an incentive to reveal her preferences truthfully because under those axioms: $A_i \succeq_F B_i$ if and only if (A_i,1/n;C,1-1/n) $\succeq_F^M$ (B_i,1/n;C,1-1/n), M from {PORnp, PORpi}. Hence both versions of POR are strongly incentive compatible for all theories that assume the reduction and independence axioms.

The above result does not imply that (either version of) POR is theoretically appropriate for testing other theories that assume reduction but do not include the independence axiom. A simple example – referred to as Example 1 in the subsequent discussion – can be used to construct a counterexample to (weak and, hence, strong as well) incentive compatibility of POR for rank dependent utility theory (RDU) by assuming the utility (given w^o) of experimental prize in the amount x is $u_{w^o}(x)=\sqrt{x}$ and the transformation of decumulative probabilities (given w^o) is the 0.9 power function. Let $V_{w^o}(L)=\int G_{w^o}^{0.9}(x)d\sqrt{x}$ be an individual's valuation of a lottery L in the experiment that pays a monetary payoff *larger* than x with probability G; the valuation represents an individual's preferences $\succeq_{w^o}$. Consider two choice options: Option A, with a sure payoff of \$30, and Option B with a 50/50 payoff of \$100 or 0. It can be easily verified that the agent with the assumed $V_{w^o}(\cdot)$ prefers Option A to Option B. Now assume the agent gets to make the choice between Option A and Option B two times and that one of the choices is randomly selected for payoff by a coin flip. Under (either version of) POR and the reduction of compound lotteries axiom, straightforward calculations reveal that Option A would be chosen in one task and Option B would be chosen in the other task because the resulting lottery {\$100, 0.25; \$30, 0.5; \$0, 0.25} in the experiment has a higher rank dependent utility, $V_{w^o}(\cdot)$ than \$30 for sure. It is true that in PORnp (unlike in PORpi) an RDU agent would not know in advance that he will be asked to choose twice between A and B but the distortion of choices is still present. The first time the subject is asked to choose between A and B he chooses A (which is truthful revelation). Having chosen A the first time, choosing B the second time is preferred to again choosing A for the same reason stated above. Therefore (either version of) POR is not incentive compatible for RDU. The same counterexample can be used to show that POR is not incentive compatible for cumulative prospect theory (Tversky and Kahneman 1992). Similarly, it can be easily verified that a dual expected utility agent (Yaari 1987) whose preferences $\succeq_{w^o}$ are represented by $V_{w^o}(L) = \int G_{w^o}^{0.9}(x)dx$ prefers a sure amount \$30 over a binary lottery that pays \$55 or 0 with 0.5 probability but with POR he prefers to choose the binary lottery in one choice and the sure amount in the other choice to choosing the sure amount two times. Therefore, POR is not incentive compatible for the dual theory.

3.3 *The Pay All Correlated (PAC and PAC/N) Mechanisms*

As shown above, the reduction and independence axioms imply that PORpi and PORnp are strongly incentive compatible. In contrast, the preference revelation properties of PAC and PAC/N depend on the *dual* independence axiom (Yaari 1987). We here show that the reduction and dual independence axioms imply that PAC and PAC/N are weakly incentive compatible for comonotonic lotteries.

For the PAC and PAC/N mechanisms, feasible events need to be defined (e.g., bingo balls numbered from 1 to 20) and all lotteries need to be arranged in the same order such that they are comonotonic. More formally, let there be given m events indexed by $s = 1,\ldots,m$ and let lotteries be identified by $A_i = (a_{i1}, p_1; \ldots; a_{im}, p_m)$ and $B_i = (b_{i1}, p_1; \ldots; b_{im}, p_m)$ where a_{is} (resp. b_{is}) is the monetary outcome of lottery A_i (resp. B_i) in state s and p_s is the probability of that state. We arrange lotteries to be comonotonic; that is, $a_{is} \geq a_{is+1}$ and $b_{is} \geq b_{is+1}$ for all s = 1,

…, m-1 and all i = 1, …, n. At the end of the experiment the state of nature is determined and, for the realized event, prizes of all chosen lotteries are paid out under PAC. Under PAC/N, the payout is 1/N of the sum of all chosen lotteries' payouts for the realized event.

As above, let an agent's choice between A_i and B_i in question i be denoted by C_i. The payoff from C_i if state of the world s occurs is denoted by c_{is}. Suppose as above that a subject made all choices apart from choice i. Then her choice between A_i and B_i will determine whether she will receive either $A_i^* = (a_{i1} + \sum_{j \neq i} c_{j1}, p_1; \ldots; a_{im} + \sum_{j \neq i} c_{jm}, p_m)$ or $B_i^* = (b_{i1} + \sum_{j \neq i} c_{j1}, p_1; \ldots; b_{im} + \sum_{j \neq i} c_{jm}, p_m)$ as reward before the state of nature is determined. This shows that PAC is weakly incentive compatible under Yaari´s (1987) dual theory: a subject whose preferences satisfy the dual independence axiom has an incentive to reveal her preferences truthfully because, under that axiom, $A_i \succeq_{w^o} B_i$ if and only if $A_i^* \succeq_{w^o}^{M} B_i^*$, M from {PAC, PAC/N}. Moreover, if lotteries are cosigned (i.e., the outcomes in a given state are all gains or all losses) PAC is also weakly incentive compatible under linear cumulative prospect theory (Schmidt and Zank 2009) since in this case the independence condition of that model has the same implications as the dual independence axiom.

Although PAC is weakly incentive compatible for the dual theory, it is not strongly incentive compatible as the following counterexample shows. Consider the A and B options in Example 1 and let the valuation of a lottery L be $V_F(L) = \int G_F^{0.9}(x)dx$, that is, the functional is linear in prizes. Assume there is background risk, F external to the experiment in which there is equal probability that wealth w will be 40 or 0. In this case, dual theory implies B $\succ_F$ A but with PAC or PAC/N the agent given two choices between Options A and B would choose Option A both times.

When we wish to compare PAC with (either version of) POR we have to keep in mind that the expected total payoff from the experiment is N times higher under PAC. This may have significant effects on behavior. In particular one may observe lower error rates under PAC as wrong decisions are more costly.[6] Therefore, we also include PAC/N in our experimental study where the payoff of PAC is divided by the number of tasks. PAC/N has the same theoretical properties as PAC; it is weakly incentive compatible under the dual theory and linear cumulative prospect theory.

It can be easily verified using the A and B options from Example 1 that PAC and PAC/N are not incentive compatible for EU or RDU. An EU agent with the (square root) utility function in Example 1 (and *no* transformation of probabilities) prefers Option A to

---

[6] Laury and Holt (2008) studies the effects of payoff scale on error rates.

Option B but with PAC (or PAC/N) the agent given two choices between Options A and B would choose Option A one time and Option B the other time. Similarly, an RDU agent with the utility *and* probability transformation functions in Example 1 prefers Option A to Option B but with PAC (or PAC/N) would make the same two choices as an EU agent.

3.4 *The Pay All Sequentially (PAS) Mechanism*

With PAS, each chosen option is paid before a subsequent decision is made. It is easy to see that PAS is not theoretically incentive compatible for the expected utility of terminal wealth (EUTW) model. We here use the (square root) utility function of experimental prizes and two options of Example 1 to illustrate possible within-experiment wealth effects with PAS for the EUTW model. An agent with the assumed square root utility function (and *no* transformation of probabilities) prefers Option A to Option B. If the agent would choose between the two options of Example 1 under PAS two times, however, the optimal strategy for the given utility function would be to choose Option B in the first choice and Option B (resp. Option A) in the second choice if the outcome of the first choice was 100 (resp. 0). Therefore, PAS is not incentive compatible for the EUTW model.[7] The possible wealth effect of PAS is not relevant to the expected utility of income model[8] or the expected utility of terminal wealth model with constant absolute risk aversion (CARA) or reference dependent preferences for which the reference point adjusts immediately after paying out the first choice (Tversky and Kahneman 1992). PAS is strongly incentive compatible for these models. PAS is weakly incentive compatible for the dual theory of expected utility (Yaari 1987)

3.5 *The Pay All Independently (PAI) Mechanism*

In the PAI mechanism, at the end of the experiment all tasks are played out independently. Theoretically, PAI has a problem, well known as portfolio effect in the finance literature: the risk of a mixture of two independent random variables is less than the risk of each variable in isolation. Due to this risk reduction effect, PAI is theoretically incentive compatible only in the case of risk neutrality. A counterexample to incentive compatibility of PAI for expected utility theory can be constructed by again using the (square root) utility function and two choice options of Example 1. The agent prefers Option A to Option B. When presenting the choice between A and B twice under PAI, however, Option B would be chosen

[7] This conclusion holds whether or not payoffs inside the experiment are integrated with payoffs outside the experiment.

[8] Three distinct expected utility models (including terminal wealth and income) are compared and contrasted in Cox and Sadiraj (2006).

both times since the resulting lottery ($200, 0.25; $100, 0.5; $0, 0.25) in the experiment has a higher expected utility. A straightforward extension shows that Example 1 provides a counterexample to incentive compatibility of PAI for RDU and CPT.

3.6 *The One Task (OT) Mechanism*

So far we can conclude that some payment mechanisms for binary choice are theoretically incentive compatible only if utility is linear in probabilities or in outcomes or if the model is defined on income rather than terminal wealth. This is not true for the OT mechanism. With this mechanism, each subject has to respond to only one choice task which is played out for real. Since there exists only one decision task, a subject has an incentive to reveal her preferences $\succeq_F$ truthfully for the most preferred option available in that task. Besides being rather costly, this mechanism has one obvious disadvantage: OT allows only for tests of hypotheses using between-subjects data. OT is nevertheless very interesting because it is the only mechanism that is always (i.e., for all possible preferences) incentive compatible.

3.7 *Summary of Incentive Compatibility Conditions*

Table 2 gives an overview of the discussion in the present section. PORpi and PORnp are strongly incentive compatible if the independence axiom holds. PAS, PAC and PAC/N are weakly incentive compatible if the dual independence axiom holds. PAS is strongly incentive

Table 2. Incentive Compatibility of Payoff Mechanisms

Preference Condition	Mechanisms
Strong Incentive Compatibility	
All theories	OT
Independence[a]	PORpi, PORnp
Income models[a]	PAS
Expected value[a]	All mechanisms
Weak Incentive Compatibility	
Dual independence[a]	PAC, PAC/N, PAS

a. Given the reduction axiom.

compatible for models defined on income rather than terminal wealth. OT is strongly incentive compatible for all theories. And, of course, all mechanisms discussed above are strongly incentive compatible for expected value theory with functional that is always linear in both payoffs and probabilities.

4. Experimental Protocol

Our first experiment includes the five pairs of lotteries reported in Table 1. Payoff in any lottery is determined by drawing a ball in the presence of the subjects from a bingo cage containing 20 balls numbered 1, 2, …, 20. Lotteries were *not* shown to participants in the format of Table 1. They were presented in a format illustrated by the example in Figure 1 which shows one of the two ways in which the lotteries of Pair 4 were presented to subjects in the experiment. Some subjects would see the Pair 4 lotteries as shown in Figure 1 while others would see them (randomly) presented with inverted top and bottom positioning and reversed A and B labeling. (See below for full details on randomized presentation of option pairs.)

<table>
<tr><td>Ball nr</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td></tr>
<tr><td>Option A</td><td colspan="5">$6</td><td colspan="15">$12</td></tr>
<tr><td>Option B</td><td>$0</td><td colspan="4">$10</td><td colspan="15">$12</td></tr>
</table>

Figure 1. An Example of Presentation of Lotteries

The experiment was run in the laboratory of the Experimental Economics Center at Georgia State University. Subject instructions are contained in an appendix available on the journal's web site. Subjects in groups OT_i, i = 1, 2, …, 5, just had to perform one choice between the lotteries of Pair i which was played out for real. Subjects in an OT_i treatment were first shown lottery Pair i at the time they made their decision. In treatment PORnp subjects were first shown a lottery pair at the time they made their decision for that pair. In all other multiple decision treatments, including PORpi, subjects were shown all five lottery pairs at the beginning of a session, as follows. Each subject was given an envelope with five (independently) randomly-ordered small sheets of paper. Each of the five small sheets of paper presented one lottery pair in the format illustrated by Figure 1. Each subject could display his or her five sheets of paper in any way desired on his or her private decision table.

Subjects entered their decisions in computers. In all treatments, including OT, the top or bottom positioning of the two lotteries in any pair and their labeling as Option A or Option B were (independently) randomly selected by the decision software for each individual subject. In all treatments other than OT, the five lottery pairs were presented to individual subjects by the decision software in independently-drawn random orders. Each decision screen contained only a single pair of lotteries.

Subjects in treatments PORpi and PORnp had to make choices for all five lottery pairs and at the end one pair was randomly selected (by drawing a ball from a bingo cage) and the chosen lottery in that pair was played out for real (by drawing a ball from another bingo cage). In treatments PAI, PAC, PAC/N, and PAS subjects had to make choices for all five pairs but here the choice from each pair was played out for real by drawing a ball from a bingo cage. In treatment PAI the five choices were played out independently at the end of the experiment whereas in treatments PAC and PAC/N the five choices were played out correlated at the end of the experiment (i.e., one ball was drawn from the bingo cage which determined the realized state of the world, hence the payoff of all five choices). In treatment PAS the chosen lotteries were played immediately after each choice was made (by drawing a ball from a bingo cage after each decision). In all treatments subjects were permitted to inspect the bingo cage and the balls before making their decisions. Each ball drawn from a bingo cage was done in the presence of the subjects and put back in the cage in the presence of the subjects.

5. Tests of Classic Properties with Data from Alternative Mechanisms

Hypothesis 1 is tested with data from each mechanism as follows. A probit model is used to estimate the probability of choosing the safer lottery in Pairs 2 and 3; right-hand variables include a dummy variable for Pair 3 and subject characteristic variables for Field (of) Study,[9] Birth Order, Female, Black, and Older than 21. The question of interest here is whether the dummy variable for Pair 3 is significantly different from 0 and, if so, the sign of the estimate will be used to determine whether our data are characterized by the fanning-in or fanning-out property. Estimates (and two-sided p-values) for all of the variables are reported in tables in the appendix. We report in the CRE column of Table 3 whether the dummy variable for Pair 3 is significantly different from zero; complete results from the probit estimation for Hypothesis 1 are reported in appendix Table A.1. We also report, in the text,

[9] Subjects were asked to report their majors. We have grouped their responses in Science and Engineering, Business and Economics and Others; the last category will be the base group in our regressions.

one-sided test results (and one-sided p-values) when there is a familiar one-sided alternative hypothesis.

First consider the test of Hypothesis 1 using data from the PAI mechanism, reported in the CRE column and last row of Table 3. We find that PAI data reject Hypothesis 1, meaning that the hypothesis of *absence of* CRE is rejected. This test result is reported as a "Yes" in the last row and first column of Table 3, which corresponds to the common practice of reporting whether a theoretical paradox "has been observed" in cases when the null hypothesis of its absence is rejected. The data provide more information. A one-sided test shows that Hypothesis 1 can be rejected at 5% significance level (one sided p-value is 0.034) in favor of fanning out of indifference curves since the coefficient for the Pair 3 dummy variable from the estimation with PAI data is negative (i.e., the safer option is chosen less often in Pair 3 than in Pair 2).[10] The specific pattern ("Fan Out") of CRE observed in the data is reported in the footnote to the "Yes" entry in the CRE column and PAI row of Table 3.

Table 3. Test Results for Hypotheses 1 - 4

Mechanism	CRE	CCE	DCRE	DCCE
OT	No	No	No	Yes[e]
PORnp	No	No	No	No
PORpi	No	Yes[b]	Yes[c]	No
PAC/N	No	Yes[b]	No	No
PAC	Yes[a]	No	Yes[d]	No
PAS	No	Yes[b]	No	No
PAI	Yes[a]	No	No	No

Notes: a. Fan Out; b. Fan In; c. IRRA; d. DRRA; e. IARA

PAC data are also consistent with the fanning out property: Hypothesis 1 is rejected in favor of the fanning out alternative hypothesis at 1% significance level (one-sided p-value is .003). This is reported in the PAC row and CRE column of Table 3 as "Yes" (meaning Hypothesis 1 is rejected, hence CRE is observed) with a footnote indicating that the CRE pattern is Fan Out.

[10] Recall that in case of a one-sided alternative hypothesis we report one-sided p-values.

Estimated coefficients for Pair 3 with data from all other mechanisms with five tasks are not significantly different from 0 (two-sided p-values $\geq$ 0.10), so all four of the multiple-choice-task mechanisms other than PAI and PAC produce data that do not reject Hypothesis 1 (absence of CRE). These findings are reported in the CRE column of Table 3 as "No", meaning CRE is not observed.

To be able to compare conclusions we draw from multiple-task treatments with those for the OT treatment we used the same method of data analysis for all treatments. Since we do not have within-subjects data for the OT treatment, we began by reporting estimates from probit regressions that use between-subjects data. But for multiple-task treatments we also have within-subjects data so we will be able to say more. Counting the number of subjects who chose the riskier option R in one of the pairs (2 or 3) and the safer option S in the other, we find the following figures (in percentages): 53% and 38% in PORnp and PORpi, 50% and 45% in PAC/N and PAC, 36% in PAS and 32% in PAI. In testing for statistical significance of these figures we need to take into account that some subjects may be indifferent between the two options within a pair. The null hypothesis that follows from the indifference argument is that frequencies of safer and riskier choices (or SR and RS patterns) are similar across Pairs 2 and 3. According to Cochran's Q test reported in the last row of Table A.1,[11] the null hypothesis of no systematic violations is rejected by PAC data (p-value is 0.008) and weakly rejected by PAI data (p-value is 0.083); these within-subjects test results are consistent with the between-subjects test results from the probit estimation.

Estimates of probit regressions using data from Pairs 3 and 4 of the probability of choosing the less risky option within a pair are used in tests of Hypothesis 2 reported in the CCE column of Table 3 (and complete results are in appendix Table A.2). The estimated coefficients for the Pair 4 dummy variable are significant for PORpi data (one-sided p-value is 0.029), PAC/N data (one-sided p-value = 0.037), and PAS data (one-sided p-value = 0.001); all of these coefficients are negative, which is consistent with indifference curves that Fan In. Estimated coefficients with data from other mechanisms are insignificantly different from 0, which is reported as "No" in the CCE column. The p-values for Cochran's Q test results reported in the last row of Table A.2 are: PORpi data (0.059), PAC/N data (0.096) and PAS data (0.007); p-vales for other mechanisms are greater than 0.1.

Data from the several mechanisms have different implications for testing expected utility theory. Five of the seven mechanisms produce data that are inconsistent with expected utility theory because the data either reject CRE or reject CCE (the entries in Table 3 are

[11] Cochran test is the same as McNemar test since we have only two groupings here.

"Yes" for presence of CRE or CEE). Furthermore, these mechanisms produce data that are variously consistent with indifference curves that Fan In, Fan Out, or are parallel.

The test results are less heterogeneous if one looks only at the three mechanisms that are theoretically incentive compatible for expected utility theory: OT, PORpi, and PORnp. Data from OT and PORnp do not reject either absence of CRE or absence of CCE (the entries in Table 3 are "No"). Data from the PORpi mechanism, however, reject absence of CCE (the Table 3 entry is "Yes") and are thus inconsistent with expected utility theory.

Results from probit tests of Hypothesis 3 that use choice data for Pairs 1 and 3 from each payoff mechanism separately are reported in the DCRE column of Table 3 (and complete results are reported in appendix Table A.3). The estimated coefficients for the Pair 3 dummy variable are insignificant with data from all mechanisms except PORpi and PAC, which is consistent with revealed risk preferences that exhibit CRRA. Estimation with data from the PAC mechanism yields a coefficient for the Pair 3 dummy variable that is significant and negative (one-sided p-value is 0.02), which is consistent with revealed risk preferences that exhibit DRRA. In contrast, estimation with data from the PORpi mechanism yields a coefficient for the Pair 3 dummy variable that is significant and positive (one-sided p-value is 0.02), which is consistent with revealed risk preferences that exhibit IRRA (for these lotteries). Results from the Cochran Q test reported in the last row of Table A.3 are generally consistent with the probit test results.

Results from probit tests of Hypothesis 4 are reported in the DCCE column of Table 3 (and complete results are reported in appendix Table A.4). Coefficients for the Pair 4 dummy variable are insignificant (two-sided p-values $\geq$ 0.10) with data from all mechanisms except OT. Revealed risk preferences with the mechanisms that involve many tasks are consistent with CARA. Estimation with OT data yields a significant coefficient that is positive (one-sided p-value is 0.012), which is consistent with preferences that exhibit IARA. The Cochran Q test results reported in the last row of Table A.4 are consistent with the probit test results.

Data from the several mechanisms have divergent implications for testing for CARA and CRRA within the range of payoffs used in the experiment. Data from three mechanisms reject either CRRA or CARA whereas data from four mechanisms do not reject either. The four mechanisms that are incentive compatible for dual theory of expected utility are OT, PAC, PAC/N and PAS. Two out of these four incentive compatible mechanisms produce data that are inconsistent with dual theory of expected utility because the data are inconsistent with either CARA or CRRA (the entries in Table 3 are "Yes" in either the DCRE or DCCE column).

We have used seven mechanisms to generate revealed risk preference data for five lottery pairs that have the potential to test for distinguishing properties of different theories of risk preferences. Out of seven mechanisms, only PORnp seems to be producing data that do not reject any of the four hypotheses. A central implication from the test results in Table 3 is that there is strong support for the view that test results for classic properties of decision theory are dependent on the payoff mechanism that is used to elicit the risk preferences.

6. Revealed Risk Preferences Differ Across Payoff Mechanisms

It has been argued in the literature (e.g., Kahneman and Tversky 1979) that subjects evaluate each choice independently of the other choice opportunities in an experiment. If this "isolation hypothesis" were to have robust empirical validity then all mechanisms in our experiment would elicit the same risk preferences. We ask whether the risk preferences revealed by subjects differ across treatments that use different payoff mechanisms or whether they are consistent with isolation of individual choices. The five columns of Table 4 present, for each lottery choice pair i (=1,2,…,5) and each elicitation mechanism, the percentage of subjects who chose the less risky (or "safer") lottery in that pair, denoted by S_i.[12] There are big differences across mechanisms in the percentages of S_i choices. Looking down the S_i columns of Table 4 we see that in three out of five columns the largest figure is more than

Table 4. Observed Frequencies (in %) of Choices of Less Risky Options

(low and high column figures in bold)

Mechanism	S_1	S_2	S_3	S_4	S_5
OT (231 subjects)	**39.47**	**15.52**	27.59	28.95	38.46
PORnp (40 subjects)	37.50	45.00	**47.50**	32.50	**60.00**
PORpi (40 subjects)	27.50	50.00	42.50	22.50	50.00
PAC/N (40 subjects)	37.50	35.00	35.00	22.50	45.00
PAC (38 subjects)	36.84	**52.63**	**23.68**	21.05	42.11
PAS (39 subjects)	**25.64**	23.08	33.33	**10.26**	**17.95**
PAI (38 subjects)	36.84	**52.63**	36.84	**34.21**	52.63

[12] The total number of subjects participating in each treatment can be found in Table 4 under the name of the treatment. The number of subjects for the one-task treatment (OT) is much larger than in all many-task treatments because of the necessity of making between-subjects inferences with OT data.

three times the smallest one: for Pair 2, choices of the safer option vary over mechanisms from 15.52% (OT) to 52.63% (PAC and PAI) or 50.00% (PORpi); for Pair 4 these figures vary from 10.26% (PAS) to 34.21% (PAI) or 32.50% (PORnp); and for Pair 5, choices of the safer option vary from 17.95% (PAS) to 60% (PORnp). The Kruskal-Wallis rank test weakly rejects at 10% significance level (chi-squared=10.97; p-value=0.089) the null hypothesis that these frequencies come from the same population.

To test for effects of mechanisms on overall revealed level of risk aversion we created a new variable, the total number of times an individual chose the less risky option. This ("Total") variable takes integer values from 0 to 5. The distributions of this overall level of risk aversion across different protocols are displayed in Figure 2. The Kruskal-Wallis rank test rejects at 1% significance level (chi-squared=16.89; p-value=0.005) the null hypothesis that observations of the variable Total observed across five-task mechanisms come from the same distribution; figures on the ranks reveal that PORnp (resp. PAS) elicits the most (resp. least) risk averse preferences.

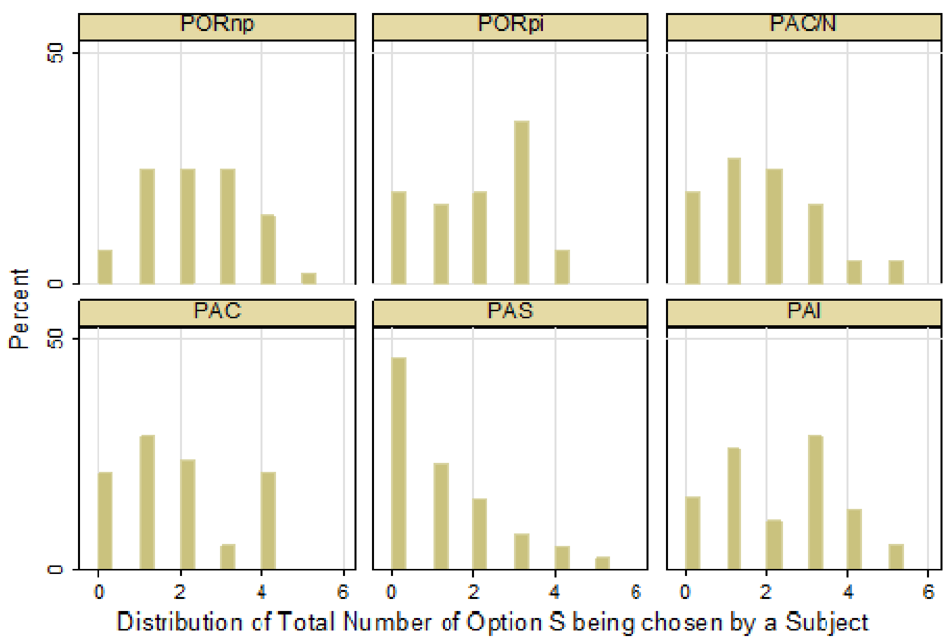

Figure 2. Distributions of Total Choices of Less Risky Options

The above tests use aggregated data. To retrieve information from data at the individual level we ran probit regressions with subject clusters to correct for correlated errors across choice tasks within an individual and with robust standard errors to accommodate heteroscedasticity.[13] Table 5 reports results from probit estimations of the probability of choosing the less risky lottery in a pair. We will discuss results from the Probit 3 column. The alternatives, Probit 1 and Probit 2 differ from Probit 3 by exclusion of some of the right-hand variables. We include these alternative specifications in the table in order to show that our central conclusions about mechanism effects are robust to alternative specifications of the estimation model.

Table 5. Probit Analysis of Choices of Less Risky Options

VARIABLES	Probit 1	Probit 2	Probit 3	Probit 3	
				Round 1	Round 5
EV Difference	-0.094 (0.360)		-0.099 (0.348)	-0.062 (0.726)	-0.108 (0.546)
VAR Difference	0.032*** (0.004)		0.034*** (0.004)	0.014 (0.518)	0.018 (0.401)
Science & Engineering		0.155 (0.110)	0.156 (0.113)	0.087 (0.566)	0.035 (0.818)
Economics & Business		0.081 (0.463)	0.085 (0.447)	0.231 (0.129)	-0.011 (0.943)
Birth Order		-0.092** (0.040)	-0.093** (0.040)	-0.124** (0.043)	-0.131** (0.034)
Female		0.309*** (0.001)	0.314*** (0.001)	0.180 (0.169)	0.292** (0.027)
Black		0.139 (0.119)	0.143 (0.115)	0.141 (0.269)	0.196 (0.127)
Older than 21		-0.153* (0.097)	-0.158* (0.091)	0.005 (0.970)	-0.164 (0.215)
DPORnp	0.445*** (0.001)	0.369*** (0.008)	0.385*** (0.007)	0.715*** (0.002)	0.489** (0.032)
DPORpi	0.288** (0.039)	0.272* (0.052)	0.286** (0.045)	0.188 (0.404)	0.533** (0.016)
DPAC	0.202 (0.187)	0.265* (0.087)	0.282* (0.074)	0.285 (0.215)	0.450** (0.048)
DPAC/N	0.196 (0.184)	0.269* (0.053)	0.286** (0.042)	0.464** (0.037)	0.066 (0.779)
DPAS	-0.193 (0.259)	-0.153 (0.370)	-0.140 (0.417)	-0.131 (0.589)	-0.200 (0.419)
DPAI	0.397*** (0.009)	0.467*** (0.002)	0.487*** (0.001)	0.438* (0.055)	0.542** (0.019)
Constant	-0.771*** (0.000)	-0.639*** (0.000)	-0.852*** (0.000)	-0.643*** (0.005)	-0.533** (0.025)
Observations (Nr. of Subjects)	1,406 (466)	1,406 (466)	1,406 (466)	466 (466)	466 (466)
Log-likelihood	-885.0	-874.7	-865.1	-286.3	-281.2

p-values in parentheses: *** p<0.01, ** p<0.05, * p<0.1

[13] Probit regressions with random effects and bootstrapped standard errors report the same results with respect to significance of the regressors that are reported in Table 5.

The right hand variables in Probit 3 include difference between expected values (EV Difference) and difference between variances (VAR Difference) of payoffs in a pair of lotteries.[14] The estimated coefficient for EV Difference is not significant.[15] The estimated coefficient for VAR Difference is significantly positive, which reveals that subjects' choices respond to differences in variance of returns, consistent with aversion to risk: the larger the variance of the riskier option relative to the safer one the more likely the safer option is to be chosen.

Some other right-hand variables are demographic controls for factors commonly associated with between-subjects differences in risk attitudes.[16] We use dummies for the subjects' field of study using three categories: Science and Engineering, Economics and Business, and Other Majors. The subject's Birth Order is significant; subjects who were an older sibling were more likely to choose the safer lottery than a younger sibling or only child. Female subjects were more likely to choose the less risky lottery. Probability of choosing the safer lottery was not significantly affected by a subject's race (Black). Being older than 21 years (weakly) affects negatively the likelihood of the less risky option being chosen.

The other variables used in the probit estimations are dummy variables for multiple decision payoff mechanism treatments. All mechanism treatment dummy variables equal 0 for OT data. Otherwise, a value equal to 1 for any one of the multiple decision payoff mechanism dummy variables selects data for that mechanism. The coefficients for all of the dummy variables for multiple decision payoff mechanisms except PAS and PAC are positive at 5% significance; PAC is positive at 8% significance. This provides support for the finding that subjects are more likely to choose the safer option (they appear to be more risk averse) with all multiple decision payoff mechanisms except PAS than they are with the OT protocol.

The PAS mechanism produces data that clearly differ from data elicited by other multiple decision mechanisms. We tested for differences between the dummy variable coefficient estimates for PAS and those for other mechanisms. Using the probit 3 model, we find that the dummy variable coefficient estimate for PAS is different from the estimates for PORnp (0.011) and PAI (0.006), and weakly different from the estimates for PAC/N (0.088)

---

[14] We included differences between EVs and VARs to control for variation in objective properties of paired lotteries that are invariant with payoff mechanisms.

[15] Differences in expected values between options within a pair were \$0.25, \$0.5 and \$2. At these small differences it is expected that this variable will have low explanatory power.

[16] Birth order has previously been reported as a significant determinant of risk attitudes (Yiannakis, 1976; Nixon, 1981; and Jobe, et al.,2006). Female subjects have previously been reported to be more risk averse than male subjects (Nixon, 1981; Jianakoplos and Bernasek, 1998; Jobe, et al., 2006; Croson and Gneezy, 2009; and Castillo, et al., 2011). Black subjects have previously been reported to be less risk averse that whites (Castillo, et al., 2011).

and PORpi (0.0999), where the figures in parentheses are Bonferroni-adjusted p-values. The (probit3) estimate for PAC is not different from the estimated coefficient for PAS; Bonferroni-adjusted p-value is 0.143.[17]

In addition to reporting an overall level of risk aversion induced by each protocol, we report in Table 6 frequencies of safer option choices, over all pairs and subjects, and the 95% confidence intervals. The percentages of the less risky option being chosen are 29%, 45%, 39%, 35%, 35%, 22% and 43% respectively, for OT, PORnp, PORpi, PAC/N, PAC, PAS and PAI data. According to these figures, PORnp and PAI seem to induce more risk averse behavior whereas OT and PAS induce less risk averse behavior. Note that the 95% confidence intervals for OT and PAS are disjoint from the ones for PORnp and PAI.

Table 6. 95% Confidence Intervals for Observed Less Risky Choice Frequencies

Mechanism	choice = S / (R+S)	[95% Conf. Interval]	Total nr of choices
OT	0.286	[0.227, 0.344]	231
PORnp	0.445	[0.376, 0.514]	200
PORpi	0.385	[0.317, 0.453]	200
PAC/N	0.350	[0.283, 0.417]	200
PAC	0.353	[0.284, 0.421]	190
PAS	0.221	[0.162, 0.279]	195
PAI	0.426	[0.355, 0.497]	190

The differences between revealed risk preferences elicited by the seven payoff mechanisms are inconsistent with the belief that subjects isolate on each decision in multiple decision experiments. The data provide support for the alternative view that the payoff mechanism chosen by the experimenter can significantly affect risk preferences revealed by the subjects. It is particularly important to note that the confidence interval for the less risky option choice frequency with PORnp is *disjoint from* the confidence intervals for OT and PAS. The inconsistency with the isolation hypothesis makes clear the importance of researching the behavioral properties of alternative mechanisms.

[17] Unadjusted p-values for PORnp and PORpi are 0.002 and 0.020, for PAC and PAC/N are 0.029 and 0.018, for PAI is 0.001.

7. Behavioral Properties of Mechanisms

What can account for the discrepancies across mechanisms in elicited risk preferences? The probit regressions reported in section 6 show that subjects were responding to the properties of lotteries within a pair. Our subjects made choices that reveal risk aversion since increase in the difference between variances of returns of the riskier and safer lottery had a positive effect on the less risky option being chosen. Other estimates from the demographics are consistent with findings in other studies. The divergent test results can be explained by failure of isolation, where such failure would be expected to cause different payoff mechanisms to elicit different risk preferences. The probit regressions reported in the right-most two columns in Table 5 for data from Round 1 and Round 5 yield further insight into the behavioral properties of the payoff mechanisms. It is important to recall that the choice order of the five lottery pairs is randomly and independently selected for each subject. Therefore Round 1 and Round 5 choices reported in Table 5 will each include a random selection of distinct lottery pairs. Hence the dummy variables for protocols in Round 1 and 5 are picking up choice order effects not lottery pair effects.

The performance of PAS shows risk preferences that are not different from OT in any comparison in Table 5, including all rounds (Probit 3) and Round 1 and Round 5. This is a particularly interesting result because, of all the multi-decision payoff mechanisms, PAS is the one that has traditionally been suspected of cross-task contamination (from wealth effects). The way in which PAS might exhibit cross-task contamination would be if there were a significant wealth effect on risk preferences, in which case risk preferences elicited in a subsequent round would not be independent of choices and outcomes in earlier rounds. Probit analysis of data from our experiment, that includes total payoff from lotteries chosen in earlier periods as an explanatory variable for choice between riskier and safer options in the current period, finds no significance of the estimated coefficient for this wealth variable: see the result reported in the Variable X coefficient row and PAS column of Table 7. This finding is consistent with earlier detailed analyses that found no significant wealth effects in other experiments that use PAS (Cox and Epstein, 1989; Cox and Grether, 1996).

In section 3 we provided some examples that illustrate the lack of incentive compatibility of mechanisms for different theories. Those examples offer insights on cross-task effects that different mechanisms might induce. We shall be testing for cross-task effects when a subject saw the tasks relevant to the hypothesis one right after the other. The example in section 3 for the PAS mechanism suggests that the payoff received in the preceding round is expected to have a negative effect on the likelihood of choosing the safer option in the current round. Probit regression reported in Table 8 using PAS data, however, reveal that the payoff in the immediately preceding round (see the Preceding Payoff (PAS) row) fails to have a significant effect on the likelihood of the safer option being chosen in the current round; the estimate is negative but insignificant (two-sided p-value = 0.106). The estimated marginal effect of payoff in the preceding round is -0.008 with a 95% confidence interval of (- 0.017, 0.002).

Table 7. Probit Tests of Cross-Task Effects

Variables / Mechanism Variable X	**PORnp (CRE)** Dummy for Pair 3	**PORnp (CCE)** Dummy for Pair 4	**PORpi (CRE)** Dummy for Pair 3	**PORpi (CCE)** Dummy for Pair 4	**PAS** Acc. Payoff	**PAI** Acc. Nr. of S choices
Variable X coefficient	0.693 (0.164)	-1.545** (0.030)	-0.171 (0.714)	-0.908* (0.061)	0.004 (0.663)	0.299** (0.030)
EV differences					-0.834*** (0.009)	0.295 (0.306)
VAR differences					0.100*** (0.002)	0.003 (0.936)
Science & Engineering	0.054 (0.914)	1.432*** (0.009)	0.396 (0.615)	0.617 (0.545)	0.302 (0.369)	-0.127 (0.708)
Economics & Business	-0.509 (0.349)	…	0.023 (0.974)	-1.741** (0.021)	0.818** (0.027)	0.155 (0.592)
Birth Order	-0.228 (0.424)	-0.330 (0.275)	-0.106 (0.805)	-0.114 (0.697)	-0.134 (0.365)	-0.186 (0.154)
Female	0.246 (0.712)	1.531* (0.080)	0.163 (0.792)	-1.766** (0.035)	0.820*** (0.006)	0.060 (0.804)
Black	0.282 (0.566)	0.157 (0.816)	-0.117 (0.820)	-1.277 (0.223)	-0.379 (0.236)	0.233 (0.342)
Older than 21	-0.115 (0.835)	-0.443 (0.648)	-0.907 (0.157)	-0.762 (0.248)	0.158 (0.587)	-0.973*** (0.000)
Preceding Payoff (PAS)					-0.029 (0.106)	
Preceding Choice (PAI)						-0.442 (0.134)
Constant	-0.492 (0.586)	-0.915 (0.259)	0.574 (0.593)	2.719** (0.039)	-1.181** (0.014)	0.428 (0.472)
Observations	32	30	32	32	195	190
Log-Likelihood	-19.14	-12.87	-20.23	-15.90	-87.88	-115.0

The entry "…" indicates that the variable was dropped because it predicted choices perfectly
p-values in parentheses: *** denotes p<0.01; ** denotes p<0.05; * denotes p<0.1

Results differ for the two implementations of POR. Consider first the highly significant, positive coefficient on the PORnp dummy variable for Round 1 reported in Table 5. In Round 1, subjects in the PORnp experimental treatment have the same lack of previous experience with lottery pair choices and the same information about lottery pairs as subjects in the OT treatment. But the highly significant positive coefficient on the Round 1 dummy variable shows that PORnp elicited much more risk averse preferences in the first round than did the OT mechanism. The only difference between these two treatments in Round 1 is that in PORnp subjects had been informed that there would be subsequent choices and that one choice would be randomly selected for payoff. This information, itself, led to much more aversion to risk in the preferences elicited in Round 1.

The alternative implementation of random selection, the PORpi mechanism, yielded quite different results in Round 1. Here, the estimated coefficient for the Round 1 dummy variable is insignificant. Recall that the difference in subjects' information across the PORnp and PORpi mechanisms at the time of a Round 1 choice consists entirely of their knowing in PORpi what the subsequent lottery choice pairs will be and their not having this information in PORnp. Together, the comparisons of PORnp with OT and PORpi data suggest that the uncertainty about future choice options that subjects faced in PORnp caused them to behave as if they were more risk averse in Round 1.

The Round 5 results look very different. Here, the dummy variable coefficients for PORnp and PORpi are almost identical. In Round 5 subjects in both treatments knew that this would be their last decision. With both versions of the random selection mechanism, the subjects were significantly more risk averse than in OT in the last round.

POR is immune to preceding-payoff cross-task effects because no lottery payoff is realized before any choice is made. In order to test for cross-task effects with POR, we test for choice order effects on revelation of classical paradoxes. In this case, as with PAS, we look at adjacent choices but now we focus on the case in which the pairs involved in a classical paradox were faced by a subject one right after the other. If there is any cross-task effect of this type one would expect it to be weaker in PORpi than in PORnp because subjects have already seen all five pairs in advance with the former implementation of the mechanism. The data support this conjecture. As shown in Table 3, PORnp does not reveal CRE or CCE when all data are used. In contrast, as shown in Table 7 (Variable X Coefficient row), if we focus only on adjacent choices then PORnp reveals a CCE (p-value = 0.030) effect but not a CRE (p-value = 0.164) effect. These results support the conclusion that PORnp data are characterized by choice order effects. For PORpi data, however, conclusions with respect to paradoxes are robust to tests with all data or tests only with adjacent round data.

Comparison of the estimated coefficients for PAC and PAC/N in Table 5 also yields behavioral insight into mechanism effects. Recall that the only difference between these two mechanisms is the scale of payoffs; experimental treatments with these two mechanisms are otherwise identical. Subjects in the PAC and PAC/N treatments have the same information about lotteries in Round 1 and Round 5 as do subjects in the PORpi treatment. Expected payoffs for PAC are N times as large as for PORpi; they are the same for PAC/N and PORpi. Choice behavior in PAC follows a similar pattern as in PORpi, with no significant difference from OT in Round 1 but (weakly) significantly more risk averse behavior by Round 5. PAC/N follows the reverse pattern, with (weakly) significantly more revealed risk aversion than OT in Round 1 but no difference from OT in Round 5.

The section 3 example of possible portfolio effects from the PAI mechanism shows how, with *uncorrelated* lotteries, a portfolio with several riskier options may be preferred to other portfolios even when the agent prefers the safer lottery to the riskier lottery in isolation. If so, then we should observe that a current choice of the safer option has a positive effect on the likelihood of the safer option being chosen latter.[18] Data are consistent with this conjecture. Probit regression reported in Table 7 (Variable X coefficient row) shows a positive effect (two-sided p-value is 0.03) of the previous total number ("Acc. Nr.") of choices of the safer option on the likelihood of choosing the safer option in the current decision task.

8. Experiment with a Hybrid Mechanism

PAS yields data that closely resemble OT but it is not incentive compatible for terminal wealth models even when they include the independence axiom. This suggests that a hybrid mechanism that combines the features of PAS and POR might have useful theoretical and behavioral properties. We here report an experiment with a mechanism in which chosen options are played out sequentially (as in PAS) before the one option relevant for payoff is randomly selected for payoff (as in POR).[19] We name this mechanism PORpas. It is incentive compatible for all theories that include the independence axiom regardless of whether they are defined on terminal wealth or on income; to verify this, note that the incentive compatibility

[18] This hypothesis is also consistent with a subject who always goes for the safe choice. But if the positive sign of the estimate of the accumulated number of less risky choices is picking up this effect then we should see a significant positive estimate in PORpi and PORnp data as well. This is not what we find; the p-values of the estimate are 0.243 and 0.493 for PORnp and PORpi.

[19] Experimenting with this hybrid mechanism was suggested by a referee. Baltussen et al. (2012) use a similar hybrid mechanism, in an experiment with the game Deal or No Deal, which includes many features not usually found in pair-wise choice experiments that could systematically affect behavior.

of PORpas can be shown with the same argument used for PORnp and PORpi in section 3.2 with the only difference being that the previous choices C_i are now replaced by the realizations of outcomes for previous choices.

We conducted an experiment with the PORpas mechanism using the five lottery pairs in Table 1. As with our PAS and PORpi treatments, the subjects were given envelopes containing five randomly-ordered small sheets of paper with the lottery pairs before making any choices. The rest of the experimental protocol is the same as reported in section 4 above.

Forty subjects participated in this experiment. We find that the mean proportion of safer choices across the five decisions is 0.265 with 95% confidence interval [0.20, 0.33]. Comparison with figures in Table 6 indicates that safer option choices in PORpas are closest to OT although also close to PAS. Looking across the five decision tasks from task 1 to task 5, the percentages of safer option choices are 40%, 25%, 15%, 25% and 27.5%.

To compare elicited risk preferences of subjects in PORpas with those in OT, PAS and the other two POR mechanisms, we use probit models of the type reported in Table 5. We find that PORpas, like PAS but unlike both PORpi and PORnp, elicits risk preferences that are not significantly different from those for OT. Details on the estimated coefficients are reported in the left-half of Table A.5 in the appendix. We find that PORpas elicits risk preferences that are not significantly different from OT for all model specifications including for all data and Round 1 and Round 5 data. Unlike data for both PORnp and PORpi in tests reported in Table 5 and in Table A.5, the PORpas data do *not* exhibit either a mechanism dummy variable effect or cross-task contamination in the form of adjacent-round effects or Round 1 or Round 5 effects. Finally, PORpas data do not exhibit significant cross-task contamination either from realized outcome effects in the immediately-preceding round or the accumulated outcomes in all preceding rounds, as reported in the right-half of Table A.5. By all these tests, data from this experiment with PORpas are free of cross-task contamination. Whether this good performance is a robust result remains to be ascertained in future research.

Finally, we use tests like those reported in Table 3 to ask whether PORpas data exhibit classic paradoxes of expected utility theory. We find that PORpas data do not exhibit either CRE or CCE. Recalling also the test results reported in Table 3, we observe that CRE is not found in data for any of the four mechanisms that are incentive compatible for expected utility theory. Out of these four mechanisms, only PORpi data exhibit CCE.

9. Comparisons to Previous Literature

Several previous studies (Camerer, 1989; Starmer and Sugden, 1991; Beattie and Loomes, 1997; Cubitt, Starmer, and Sugden, 1998; Hey and Lee, 2005a, 2005b; and Lee, 2008) tested whether POR elicits true preferences and often reported the conclusion that serious distortions were not observed. But many of the authors' stated conclusions are based on experimental protocols or tests of hypotheses that do not actually support the conclusion of robust absence of preference distortion by POR. In this section we will discuss some differences between our findings, and the experimental protocols and tested hypotheses of other studies, which can account for what may only appear to be inconsistent empirical findings.

Camerer (1989) tests the incentive compatibility of POR by giving subjects the possibility to change their choices after the question relevant for payoff has been selected. Since only very few people do so, he concludes that POR does not seem to induce biases. This conclusion, however, implicitly relies on the assumption that decision makers are "naïve" in the terminology of Machina (1989); such decision makers will be dynamically inconsistent. In contrast, if decision makers are "resolute" (Machina, 1989) the other options involved in the POR mechanism could lead to altered preferences *and these altered preferences would still hold after selection of the choice problem relevant for payoff*, which would cause subjects to stick with initially-biased preference revelation.

Starmer and Sugden (1991) and Hey and Lee (2005a)[20] test the isolation hypothesis against the alternative hypothesis of "full reduction". The isolation hypothesis is that subjects consider each choice task in the experiment in isolation from all of the others. The alternative, full reduction hypothesis is that subjects consider all choice tasks in the experiment as being part of one big lottery and make all choices so as to yield the most preferred probability distribution of payoffs from the whole experiment. Tests in these two papers reject the hypothesis of full reduction in favor of the alternative of isolation. But there are many alternatives to isolation other than full reduction, which is *a priori* implausible in experiments in which subjects make a large number of choices and first see a lottery pair when they are asked to make the choice for that pair (as noted by Hey and Lee, 2005a).

[20] In the experiment reported in Hey and Lee (2005a, 2005b), one out of 179 subjects was selected to receive payment for one out of his or her 30 choices.

Hey and Lee (2005b) report a test that uses a "partial reduction" hypothesis as an alternative to the isolation hypothesis. The partial reduction hypothesis is that subjects consider all preceding choice tasks and the current choice task in the experiment as being part of one big lottery and make each current choice so as to yield the a most preferred probability distribution of payoffs over all choices through the current one. They consider two versions of the partial reduction hypothesis which differ according to whether the current choice task is given the same weight or higher weight than preceding choice tasks. They report that isolation appears to explain the data better than either form of partial reduction.

Both full reduction and partial reduction appear to require information processing that could exceed subjects' capacities or willingness in all but the simplest experiments with few choices. In that way, these are extreme alternatives to the isolation hypothesis. More plausible alternatives to isolation are provided by hypotheses about cross-task contamination in which a subject's answer in one decision task may be affected by the content of some other decision tasks or the choices made in some adjacent tasks. As reported above, POR data from our experiment exhibit significant cross-task contamination, which is *inconsistent* with the isolation hypothesis. This leads us to ask whether there is evidence of cross-task contamination in previous experiments with POR.

9.1 Cross-Task Contamination from POR in Previous Experiments

Starmer and Sugden (1991) recognized the possibility of cross-task contamination from POR and reported a test for one form it might take. They report (on pg. 977) that their two-tailed test for cross-task contamination in their POR data was marginally significant (the reported p-value is 5.1%).[21] Starmer and Sugden use POR and an "impure" form of OT in which subjects make one choice that pays money following many hypothetical choices. Beattie and Loomes (1997) employ "pure" OT in which the one task that pays money is not embedded in other decision tasks, that is, a subject actually makes only one choice (as in our experiment reported above). They find a significant difference between responses to POR and OT in one of four analyzed choice problems. This result is consistent with results we report for our implementation of pure OT in the experiment reported in preceding sections of our paper; the results are inconsistent with isolation. Lee (2008) compares responses under POR and PAS and finds evidence for decreasing absolute risk aversion (i.e., less risk aversion

[21] This result seems to be consistent with our finding of cross-task contamination by POR unless one insists: (a) on a specific *two*-tailed test; and (b) that a p-value of 5.1% rather than 5.0% leads to opposite conclusions. Starmer and Sugden (1991, pg. 977) state (in our view correctly) that "… we cannot claim to have proved, on the basis of such a test, that the random-lottery incentive system is unbiased."

under PAS, which is a mechanism effect on revealed risk preferences). Such mechanism effects are inconsistent with isolation.

9.2 Experiment with Impure OT

We conducted a new OT treatment using a payoff protocol similar to the one in Starmer and Sugden (1991) and Cubitt, Starmer and Sugden (1998).[22] The implementation of the OT mechanism in those papers was "impure" in that the one choice task that was paid followed a series of hypothetical choices. We use a protocol similar to Starmer and Sugden (1991) in which OT follows a series of hypothetical choices over pairs of lotteries that are all shown to the subjects before any choices are made. Pairs of lotteries in the new treatment are the same as the five pairs used in other treatments reported in this paper (see Table 1 for details).

Seventy-seven subjects participated in this experiment. Subjects were given envelopes with five lottery pairs in random order and could open the envelopes and see the contents before making any decisions. This part of the protocol is the same as used in all other treatments except PORnp, in which a subject would first see an option pair at the time of making a decision for that pair. The first four decision tasks have hypothetical payoffs. The fifth task is paid for sure. We analyze data from the fifth task. In that task, 26 subjects were given option Pair 2, 26 subjects were given Pair 3, and 25 subjects were given Pair 4. Each subject was given four other option pairs in independent random order.[23]

Based on findings reported in section 6, we hypothesized that embedding the one paid round in Impure OT in a multiple decision treatment (with four hypothetical payoff decisions) would have an effect on elicited risk preferences similar to that in PORpi and PORnp: that it would increase the proportion of safer option choices. This is what we find: the percentage of safer choices for the three pairs (2, 3, and 4) is 23.4% with OT and 35% with Impure OT. To compare the elicited risk preferences of subjects in Impure OT with those in the other treatments, we estimate probit models of the type reported in Table 5 but now use only data for tasks 2, 3, and 4 because they are the ones that are paid in the Impure OT treatment. These probit results are reported in appendix Table A.6. The Probit 3 estimates reveal that risk preferences elicited by Impure OT are significantly more risk averse than those for OT.

We next ask whether CRE or CCE is observed with Impure OT data. We find that in the paid round, the safer option was chosen by 30.77%, 42.31% and 32% of subjects in pairs

[22] This experiment was suggested by a referee.
[23] We select only CRE and CCE lottery pairs for payoff because that is what Starmer and Sugden (1991) did.

2, 3 and 4. Data show neither CRE nor CCE: Fisher's exact test reports a p-value of 0.565 for both. Similar to the across-subjects data analysis for OT (reported in the first columns of Tables A.1 and A.2 in the appendix), we also ran probit regressions using Pair 2, 3, and 4 data and a dummy variable for Pair 3 data (see table A.6, CRE and CCE coumns). Again, the Pair 3 estimated coefficient is not significant; the two-sided p-values are 0.282 and 0.287. We conclude that Hypotheses 1 and 2 are not rejected by our Impure OT data; neither CRE nor CCE is observed.

10. Implications for Experiments on Social Preferences

Research on the properties of payoff mechanisms has implications for many topics other than lottery pair choices. One illustrative example is provided by experiments on social preferences. Many experiments on social preferences involve decisions under risk or uncertainty. For example, if a first mover in an investment game experiment (Berg, Dickhaut, and McCabe 1995) sends money to the second mover the first mover's monetary payoff in the experiment is risky because it depends on the subsequent return decision of the second mover. Many social preferences experiments involve one shot games, hence are using OT. But many other experiments include multiple decisions under risk (or uncertainty) and use some mechanism for paying the subjects. The most commonly used mechanism for multiple-decision social preferences experiments is PAS (e.g. Bohnet, Grieg, Hermann, and Zeckhauser 2008, Charness and Haruvy 2002, Fehr and Gachter 2000, 2002, and Fehr and Schmidt 2004), which has internal theoretical validity for income models of decision under risk. Many papers use POR (e.g. Andreoni and Miller 2002, Ashraf, Bohnet, and Piankov 2006, Goeree, Holt, and Laury 2002, and Vanberg 2008), which has internal theoretical validity for expected utility theory. Other papers use PAI (e.g., Burks, Carpenter, and Verhoogen 2003, Chaudhuri and Gangadharan 2007), which has internal theoretical validity only for expected value theory; for risk averse subjects, these experiments involve portfolio incentives that confound drawing conclusions from the data.

Experiments intended to "identify" trust and reciprocity or "decompose" trust and trustworthiness inherently involve decisions in more than one game or decision task. Some papers (e.g. Cox 2004) use between-subjects designs and OT. Other papers (e.g., Ashraf, Bohnet, and Piankov 2006) use within-subjects designs and PORnp. The within-subjects design with PORnp has internal theoretical validity for expected utility theory but there can be cross-task contamination that biases data with this type of protocol, as reported in papers that tested for it. For example, Cox (2009) reports an experiment in which data show that informing subjects there will be another unspecified decision task following a dictator game significantly shifts their behavior towards greater generosity in an experiment in which there is anonymity (because of double-blind payoffs) and random selection of one task for payoff. With this implementation of the PORnp mechanism, subjects do not isolate their play in a dictator game from the other decision task in the experiment. Cox, Sadiraj, and Sadiraj (2008) report three experiments with different designs for the moonlighting game and dictator control games. Experiment 2 has a within-subjects design for the moonlighting game and dictator control games and uses the PORnp mechanism to pay subjects. Experiment 3 has a between-subjects design for the moonlighting game and dictator control game and pays subjects with OT. Data show that subjects are more trusting and fearful in Experiment 2 than in Experiment 3. With this implementation of the PORnp mechanism, subjects do not isolate their play in the moonlighting game from their play in the dictator game. Both of these papers (Cox 2009 and Cox, Sadiraj, and Sadiraj 2008) test for and find cross-task contamination from the PORnp mechanism in the context of social preferences experiments.

Finally, Sadiraj and Sun (2012) conduct an experiment on bargaining with alternating offers on gain and loss domains using POR and Impure OT payoff protocols. They report that Impure OT induces more efficient bargaining behavior than POR and that the effect is more pronounced when subjects bargain over the distribution of gains than the shares of losses. These findings are inconsistent with isolation.

11. Implications for Choice of Mechanism

If subjects generally were to isolate on each decision they make within a multiple-decision context then the choice of mechanism would be unimportant detail of experimental protocols. Data from our experiments are inconsistent with the isolation hypothesis. This makes clear the central importance of appropriate choice of mechanism to validity of conclusions drawn from data. There are two distinct questions that arise in evaluating mechanisms: (a) incentive compatibility and (b) behavioral bias. Incentive compatibility is a straightforward logical question. Mechanism behavioral bias is a more subtle empirical question. Table 2 and section 8 report incentive compatibility properties of payoff mechanisms. Sections 5 through 10 report test results for behavioral properties of mechanisms. This leads us to the topic of spelling out implications of our theoretical and empirical analysis for experimental methods. We consider three ways of looking at this issue that differ in terms of the objectives of particular applications of experimental methods.

11.1 All or Nothing Approach to Testing a Theory

One approach to testing hypotheses that follow from a particular theoretical model is to use a payoff mechanism that is incentive compatible for that theory, test the hypotheses, and state conclusions about the theory. Consider, for example, tests of CRE or CCE. These are hypotheses that follow from the axioms of expected utility theory and those same axioms imply that POR is incentive compatible. Hence a test for CRE or CCE that uses PORnp or PORpi or PORpas has internal theoretical validity. Analogously, tests of DCRE or DCCE that use PAC or PAC/N or PAS have internal theoretical validity because those mechanisms are incentive compatible for the dual theory of expected utility (Yaari 1987). PAC and PAC/N are also incentive compatible for linear cumulative prospect theory (Schmidt and Zank 2009). PAS is incentive compatible for testing the expected utility of terminal wealth model with constant absolute risk aversion. PAS is incentive compatible for testing any theory defined on income, rather than terminal wealth, such as dual theory of expected utility, the expected utility of income model (Cox and Sadiraj 2006), and reference dependent preferences for which the reference point adjusts immediately after paying out the first choice (Tversky and Kahneman 1992).

PAI is not incentive compatible for any of the theoretical models of risk aversion considered in this paper and hence does not produce data for any tests with internal theoretical validity except for expected value theory. In contrast, OT is incentive compatible for all theories considered herein and hence appropriate for testing hypotheses derived from any of them.

One puzzle is provided by the widespread use of POR rather than PAS to test hypotheses for cumulative prospect theory (CPT); see, for examples, papers by Birnbaum (2004, 2008), Kothiyal, Spinu and Wakker (2013), Harrison and Rutström (2009), and Wakker, Kobberling, and Schwieren (2007). It was generally known after results in Holt (1986) that POR places crucial reliance on the independence axiom that was subsequently explicitly discarded under CPT (Tversky and Kahneman, (1992)), which makes POR inappropriate for tests of CPT with internal theoretical validity. Furthermore, CPT was also specifically developed as a model defined on income, not terminal wealth, hence wealth effects are not relevant. This means that PAS is a mechanism that could have been used in tests of CPT that *would* have had internal theoretical validity.[24] Authors who did use PAS in tests of CPT include Kachelmeir and Shehata (1992).

11.2 Nuanced Approach to Testing a Theory

There are issues distinct from incentive compatibility that arise in a nuanced approach to testing theory in which the researcher is concerned about the source of consistency or inconsistency with hypotheses. A good example is provided by the tests of CRE and CCE reported in sections 5, 8 and 9. PORpi is incentive compatible for expected utility theory (EUT), hence the significant inconsistency with CCE with data from that mechanism has internal theoretical validity. But PORnp, PORpas, OT, and Impure OT are also incentive compatible for EUT and data from our treatments with those mechanisms do not exhibit CCE. The difference in test results comes from the different behavioral properties of the payoff mechanisms, all of which are incentive compatible for EUT. A nuanced approach to testing a theoretical hypothesis will try to discriminate between inconsistencies with theory that are specific to one incentive compatible payoff mechanism and patterns of inconsistency that are robust to other incentive compatible mechanisms.

11.3 Discriminating Between Theories

Research on decision under risk includes experiments designed to discriminate between alternative theories; see, for examples, Camerer (1989) and Hey and Orme (1994). Design of experiments of this type encounters an especially difficult issue of incentive compatibility because a payoff mechanism that is incentive compatible for one of the theoretical models being compared is typically not incentive compatible for one or more of the other theoretical models if subjects make multiple decisions. This problem is present in the experiments reported by Camerer (1989) and Hey and Orme (1994) that asked subjects to make multiple decisions and paid them using some version of POR. Such experiments could be conducted using OT, as that mechanism is an incentive compatible mechanism for all theories being compared. Experiments comparing cumulative prospect theory with the expected utility of income model would have theoretical validity if they used PAS because both models are defined on income, not terminal wealth. Experiments comparing linear cumulative prospect theory with dual theory of expected utility using cosigned lotteries would have theoretical validity if they used PAC, PAC/N or PAS.

[24] The literature on CPT experiments also includes many papers in which subjects were not paid salient rewards for any decision (e.g, Abdellaoui, Barrios, and Wakker 2007, Birnbaum and Chavez 1997, Bleichrodt, Pinto, and Wakker 2001, Lopes and Oden 1999, and Gonzales and Wu 1999).

We now ask whether data from a mechanism other than OT can be used to discriminate between alternative theories. After all, data reveal subjects' preferences over options conditional on the incentive properties of the mechanism. Consider, for example, PORpi and three decision tasks of choosing between options A_i and B_i for i=1,2,3. Regardless of which theory the researcher is applying, observations such as choice of A_1 (rather than B_1) and B_2 (rather than A_2) and A_3 (rather than B_3) with PORpi payoffs do *not* tell us how the subject ranks, say, lotteries P = (B_1,1/3;B_2,1/3;A_3,1/3) or Q = (A_1,1/3;B_2,1/3;B_3,1/3). Observed choices of A_1 and B_2 and A_3 would, however, reveal that the subject ranks lottery R = (A_1,1/3;B_2,1/3;A_3,1/3) weakly higher than any other feasible lottery, including P and Q. Hence, it is (theoretically) correct to conclude from the observed choices that a theory's functional of lottery R has higher value than the same theory's functional of P (or Q, or another feasible compound lottery), for any theory that the researcher has chosen as a maintained hypothesis. One can use this information to infer something about risk preferences of the subject, given the theory, by using the composition of the theory's functional and the functional form of the payoff mechanism.[25] The problem with many studies that test non-EU theories with POR data is that they interpret the observed choices (A_1 and B_2 and A_3) as revealing that the subject ranks option A_1 weakly higher than option B_1, and B_2 weakly higher than A_2, and A_3 weakly higher than B; this is not correct unless the maintained theory is EU (see section 3).

12. Summary and Conclusion

Experiments on choice under risk typically involve multiple decisions by individual subjects and use of a payoff mechanism to implement incentive payoffs. If subjects were to isolate each individual decision from other decisions then choice of payoff mechanism would be an unimportant detail of experimental protocols. Our data are inconsistent with the hypothesis that subjects' revealed risk preferences are isolated from payoff mechanism effects. This failure of isolation in the context of binary lottery choice experiments reported herein, together with previously reported failures of isolation in social preferences experiments using the PORnp mechanism, raise serious issues for experimental methods.

We have reported data in this paper from experimental treatments with nine payoff mechanisms, some of which are strongly incentive compatible for expected utility theory, some weakly incentive compatible for dual theory of expected utility, one strongly incentive

[25] This approach is used by Harrison and Swarthout (2012) and Harrison, Martínez-Correa, and Swarthout (2013) in papers in which OT is used together with PORnp to inform structural estimation of models of risk preferences.

compatible for cumulative prospect theory, and one incentive compatible only for expected value theory. Our data reveal that different mechanisms elicit data that have different implications for fundamental properties of decision theories such as the independence axiom vs. fanning in or fanning out as well as the dual independence axiom vs. increasing or decreasing relative or absolute risk aversion.

Five of the nine mechanisms are strongly incentive compatible for expected utility theory; they are OT, Impure OT, PORnp, PORpi, and PORpas. A remarkable result from our experiments is the near absence of evidence of paradoxes for expected utility theory with data from the incentive compatible mechanisms. None of these mechanisms elicited data that exhibit CRE. Only one (PORpi) out of these five mechanisms elicited data that exhibit CCE. These results call into question whether there is robust empirical support for Allais paradox inconsistencies with expected utility theory. Recall our distinction in section 11 between the "all or nothing" and "nuanced" approaches to testing theoretical hypotheses. If some observed Allais patterns in data result from behavioral consequences of using only one specific payoff mechanism, and are not a feature of risk preferences that is robust to other incentive mechanisms, then implications for theory are not compelling (at least for a nuanced approach to empirical research).

Although PORnp and PORpi are strongly incentive compatible for testing hypotheses from expected utility theory, the changes in elicited risk preferences across rounds in our experiment (see Table 5) and the adjacent-decision cross-task effects (see Table 7) raise serious questions about the behavioral properties of these two alternative implementations of the random decision selection mechanism. In contrast, PAS elicited risk preferences that did not vary between rounds. This reflects the absence of significant wealth effects found in this study and in two previous studies (Cox and Epstein 1989, Cox and Grether 1996) that carefully analyzed PAS data for wealth effects. Our PAS data also do *not* exhibit a significant preceding-round outcome effect nor a cumulative outcome effect on current option choice. Unlike PAS, the hybrid PORpas mechanism is strongly incentive compatible for all expected utility models including the expected utility of terminal wealth model. Data from the PORpas mechanism are also free from adjacent-decision cross-task contamination and free from preceding outcome and accumulated outcome effects. Data from PAS and PORpas are not significantly different from data for OT.

Empirical failure of isolation from mechanism effects can be especially a problem for design of experiments to test theories such as rank dependent theories and betweenness theories that do not include either the independence axiom or the dual independence axiom.

When such models are defined on terminal wealth, there is no known payoff mechanism that is incentive compatible for *multiple* decision experiments. In contrast, PAS provides an incentive compatible mechanism for use in testing models defined on income in which the reference point adjusts to zero income after each payoff realization; this holds for both rank dependent (Tversky and Kahneman 1992) and expected utility (Cox and Sadiraj 2006) income models.

The theoretical rationale for use of POR rather than PAS in multiple decision experiments with the expected utility of terminal wealth (EUTW) model is, of course, to control for possible wealth effects on sequential decisions. This suggests questions about plausibility of modeling risk aversion over small-scale risks with concave utility (Rabin 2000, Cox and Sadiraj 2006, Neilson 2001, and Safra and Siegel 2008). Regardless, however, of whether concave utility is abandoned, modeling risk aversion through nonlinearity of probabilities has its own calibration problems and therefore its plausibility is also questioned (Sadiraj, forthcoming). Experimental data that support empirical relevance of both (nonlinear payoff transformation and nonlinear probability transformation) calibration critiques already exist (Cox, Sadiraj, Vogt and Dasgupta, in press). Recent developments of third generation prospect theory (e.g., Schmidt, Starmer and Sugden 2008) try to avoid calibration critique with reference dependence that can include reference lotteries rather than reference amounts of money. Questions arise about how to test such models that are beyond the scope of the present paper. Within our scope is the observation that there does not yet exist an incentive compatible mechanism for use in multiple decision experiments with this theory; until such a mechanism is designed any attempts to test it with theoretically unbiased risk preference elicitations would need to use OT.

The OT mechanism avoids any possible cross-task contamination and is the only known incentive compatible payoff protocol to use in experiments designed to test all theories. OT, however, does have limitations in that it is expensive to use in experiments and it requires that all hypothesis tests of choice patterns such as CRE and CCE be conducted between subjects.

The finding that experimental data are inconsistent with the isolation hypothesis makes clear the importance of systematic study of the properties of alternative payoff mechanisms and the relationship of those properties to validity of conclusions about theory that can be drawn from data. Our work offers a step in the direction of such study.

References

Abdellaoui, M., Barrios, C. and Wakker, P. P. (2007). "Reconciling Introspective Utility with Revealed Preference: Experimental Arguments Based on Prospect Theory", *Journal of Econometrics* 138.1, pages 356-378.

Allais, M. (1953), "Le Comportement de l'homme Rationnel devant le risque: Critique des Postulats et Axiomes de l'école Américaine", *Econometrica*, 21, 503-546.

Andreoni, James & John Miller, 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism", *Econometrica*, 70(2), 737-753.

Ashraf, Nava, Iris Bohnet, Nikita Piankov, 2006. "Decomposing Trust and Trustworthiness", *Experimental Economics, Special Issue on Behavioral Economics*, 9(3), 193-208.

Azrieli, Y., Chambers, C. P. and Healy, P. J. et. al. (2012). "Eliciting Multiple Choices: A Theoretical Analysis of Incentives in Experiments," Working Paper, Ohio State University.

Baltussen, Van den Assen, M., Post, T. and Wakker, P. (2012), "Random Incentive Systems in a Dynamic Choice Environment", *Experimental Economics*, 15, 418-443.

Beattie, J. and Loomes, G. (1997), "The Impact of Incentives upon Risky Choice Experiments", *Journal of Risk and Uncertainty*, 14, 149-62.

Berg, J., Dickhaut, J., McCabe, K. (1995). "Trust, Reciprocity, and Social History", *Games and Economic Behavior*, 10, 122–142.

Birnbaum, M. H. (2004). "Tests of Rank-dependent Utility and Cumulative Prospect Theory in Gambles Represented by Natural Frequencies: Effects of Format, Event Framing, and Branch Splitting", *Organizational Behavior and Human Decision Processes*, 95, 40–65.

Birnbaum, M. H. (2008). "New Tests of Cumulative Prospect Theory and the Priority Heuristic: Probability-outcome Tradeoff with Branch Splitting", *Judgment and Decision Making*, 3(4), 304–316

Birnbaum, M. H. and Chavez, A. (1997). "Tests of Theories of Decision Making: Violations of Branch Independence and Distribution Independence", *Organizational Behavior and Human Decision Processes*, 71(2), 161–194.

Bleichrodt, H., Pinto, J. L. and Wakker, P. P. (2001). "Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility. *Management Science* 47.11, 1498-1514

Bohnet, I., Greig, F., Herrmann, B. and Zeckhauser, R. (2008). "Betrayal Aversion." *American Economic Review*, 98(1), 294-310.

Burks, S. V., Carpenter, J. P. and Verhoogen, E. (2003). "Playing Both Roles in the Trust Game", *Journal of Economic Behavior and Organization*, 51, 195-216.

Camerer, C. F. (1989), "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty*, 2, 61-104.

Castillo, M., Ferraro, P., Jordan, J. L. and Petrie, R. (2011). "The Today and Tomorrow of Kids:Time Preferences and Educational Outcomes of Children", *Journal of Public Economics*, 95 (11-12), 1377-1385.

Charness, G. and Haruvy, E. (2002). "Altruism, equity, and Reciprocity in a Gift-exchange Experiment: An Encompassing Approach", *Games and Economic Behavior*, 40(2), 203-231.

Chaudhuri, A. and Gangadharan, L. (2007). "An Experimental Analysis of Trust and Trustworthiness", *Southern Economic Journal*, 73, 959-985.

Cox, J. C. (2004). "How to Identify Trust and Reciprocity," *Games and Economic Behavior,* 46(2), 260-281.

Cox, J. C. (2009). "Trust and Reciprocity: Implications of Game Triads and Social Contexts," *New Zealand Economic Papers. Special Issue: Laboratory Experiments in Economics, Finance and Political Science*, 43(2), 89 – 104.

Cox, J. C. and Epstein, S. (1989). "Preference Reversals Without the Independence Axiom", *American Economic Review*, 79(3), 408-426.

Cox, J. C. and Grether, D. A. (1996). "The Preference Reversal Phenomenon: Response Mode, Markets, and Incentives", *Economic Theory*, 7(3), 381-405.

Cox, J. C. and Sadiraj, V. (2006). "Small- and Large-Stakes Risk Aversion: Implications of Concavity Calibration for Decision Theory", *Games and Economic Behavior*, 56, 45-60.

Cox, J. C., Sadiraj, K. and Sadiraj, V. (2008). "Implications of Trust, Fear, and Reciprocity for Modeling Economic Behavior", *Experimental Economics*, 11, 1-24.

Cox, J .C., Sadiraj,V., Vogt, B. and Dasgupta, U. Economic Theory (in press). "Is There A Plausible Theory for Decision under Risk? A Dual Calibration Critique", *Economic Theory*. Advance online publication. doi 10.1007/s00199-012-0712-4

Croson, R. and Gneezy, U. (2009). "Gender Differences in Preferences,"*Journal of Economic Literature*, 47, 448-474.

Cubitt, R. P., Starmer, C. and Sugden, R. (1998). "On the Validity of the Random Lottery Incentive System", *Experimental Economics*, 1, 115-131.

Cubitt, R. P., Starmer, C., and Sugden, R. (2001). "Discovered Preferences and the Experimental Evidence of Violations of Expected Utility Theory," *Journal of Economic Methodology*, 8, 385-414.

Fehr, E. and Gächter, S. (2000). "Cooperation and Punishment in Public Goods Experiments", *American Economic Review*, 90(4), 980-994.

Fehr, E. and Gächter, S. (2002). "Altruistic Punishment in Humans", *Nature*, 415, 137-140.

Fehr, E. and Schmidt, K. M. (2004). "Fairness and Incentives in a Multi-task Principal-Agent Model", *Scandinavian Journal of Economics*, 106(3), 453-474.

Goeree, J. K. and Holt, C. A. and Laury, S. K. (2002). "Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior", *Journal of Public Economics*, 83(2), 255-276.

Gonzalez, R. and Wu, G. (1999). "On the Shape of the Probability Weighting Function", *Cognitive Psychology*, 38, 129–166

Harrison, G. W., Martínez-Correa, J. and Swarthout, J. T. (2013). "Reduction of Compound Lotteries with Objective Probabilities: Theory and Evidence", Working Paper, Center for the Economic Analysis of Risk, Georgia State University.

Harrison, G.W. and Rutström, E. E. (2009). "Expected Utility and Prospect Theory: One Wedding and a Decent Funeral", *Experimental Economics*, 12, 133-1587.

Harrison, G.W. and Swarthout, J. T. (2012). "The Independence Axiom and the Bipolar Behaviorist", Experimental Economics Center Working Paper 2012-01, Georgia State University.

Hey, J. D. and Lee, J. (2005a). "Do Subjects Separate (or Are They Sophisticated?", *Experimental Economics*, 8, 233-265.

Hey, J. D. and Lee, J. (2005b). "Do Subjects Remember the Past?", *Applied Economics*, 37, 9-18.

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

Holt, C. A. (1986). "Preference Reversals and the Independence Axiom", *American Economic Review*, 76, 508-515.

Holt, C. A. and Laury, S. K. (2002). "Risk Aversion and Incentive Effects", *American Economic Review*, 92, 1644-1655.

Jianakoplos, N. A. and Bernasek, A. (1998). "Are Women More Risk Averse?", *Economic Inquiry*, 36, 620 – 630.

Jobe, J. B., Holgate, S. H. and Scrapansky, T. A. (2006). "Risk Taking as Motivation for Volunteering for Hazardous Experiment, *Journal of Personality*, 51 (1), 95-107.

Kachelmeier, S. J. and Shehata, M. (1992). "Examining Risk Preferences under High Monetary Incentives: Experimental Evidence from the People's Republic of China" *American Economic Review*, 82(5), 1120-1141.

Kahneman, D. and Tversky, A. (1979), "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47, 263-291.

Kothiyal, A., Spinu,V. and Wakker, P. (2013). An Experimental Test of Prospect Theory for Predicting Choice under Ambiguity. *Journal of Risk and Uncertainty*, forthcoming.

Laury, S. K. and Holt, C. A. (2008). "Payoff Scale Effects and Risk Preference under Real and Hypothetical Conditions", in C. Plott and V. Smith (eds.) *Handbook of Experimental Economic Results*, Vol. 1, New York, Elsevier Press.

Lee, J. (2008), "The Effect of the Background Risk in a Simple Chance Improving Decision Model", *Journal of Risk and Uncertainty*, 36, 19–41.

Lopes, L. L. and Oden, G. C. (1999). "The Role of Aspiration Level in Risky Choice: A Comparison of Cumulative Prospect Theory and SPA Theory", *Journal of Mathematical Psychology* 43, 286-313.

Machina, M. J. (1982). "'Expected Utility' Analysis without the Independence Axiom", *Econometrica*, 50, 277-323.

Machina, M. J. (1989). "Dynamic Consistency and Non-expected Utility Models of Choice under Uncertainty," *Journal of Economic Literature,* 27(4), 1622-1668.

Neilson, W. S. (1992). "A Mixed Fan Hypothesis and its Implications for Behavior Toward Risk", *Journal of Economic Behavior and Organization*, 19, 197-211.

Neilson, W. S. (2001). "Calibration Results for Rank-Dependent Expected Utility", *Economics Bulletin*, 4, 1-5.

Nixon H. L. (1981). "Birth Order and Preferences for Risky Sports among College Students," *Journal of Sport Behavior*, 4, 12-23.

Rabin, M. (2000). "Risk Aversion and Expected Utility Theory: A Calibration Theorem", *Econometrica*, 68, 1281-1292.

Sadiraj, V. (forthcoming). "Probabilistic Risk Attitudes and Local Risk Aversion: A Paradox", *Theory and Decision*.

Sadiraj, V. and Sun, J. (2012). "Efficiency in Bargaining Games with Alternating Offers", *Economics Bulletin*, 32 (3), 2366-2374.

Safra, Z. and Segal, U. (2008). "Calibration Results for Non-Expected Utility Theories", *Econometrica*, 76, 1143-1166.

Safra, Z. and Segal, U. (2009). "Risk Aversion in the Small and in the Large: Calibration Results for Betweenness Functionals", *Journal of Risk and Uncertainty*, 38, 27-37.

Schmidt, U., Starmer, C. and Sugden, R. (2008). "Third Generation Prospect Theory", *Journal of Risk and Uncertainty*, 36, 203-223.

Schmidt, U., and Zank, H. (2009), "A Simple Model of Cumulative Prospect Theory", *Journal of Mathematical Economics,* 45, 308–319.

Starmer, C. and Sugden, R. (1991), "Does the Random Lottery Incentive System Elicit True Preferences? An Experimental Investigation", *American Economic Review*, 81, 971–978.

Tversky, A., Kahneman, D. (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty*", Journal of Risk and Uncertainty,* 5, 297-323.

Vanberg, C. (2008). "Why Do People Keep Their Promises? An Experimental Test of Two Explanations -super-1," *Econometrica*, 76(6), 1467-1480.

Wakker, P. P., Köbberling, V. and Schwieren, C. (2007). Prospect-Theory's Diminishing Sensitivity versus Economics' Intrinsic Utility of Money: How the Introduction of the Euro Can be Used to Disentangle the Two Empirically", *Theory and Decision*, 63(3), 205-231.

Yaari, M. E. (1987), "The Dual Theory of Choice under Risk", *Econometrica*, 55, 95-115.

Yiannakis A. (1976). "Birth Order and Preference for Dangerous Sports among Males," *Research Quarterly*, 47, 62-67.

Appendix Tables

Table A.1. Tests of Hypothesis 1 (CRE)

Variables / Mechanism	**OT**	**PORnp**	**PORpi**	**PAC/N**	**PAC**	**PAS**	**PAI**
Probit Regression							
Pair 3 (D)	0.424 (0.135)	0.055 (0.859)	-0.206 (0.429)	0.007 (0.984)	-0.844*** (0.006)	0.323 (0.310)	-0.498* (0.068)
Science & Engineering	0.036 (0.914)	0.292 (0.367)	0.441 (0.310)	0.520 (0.111)	0.488 (0.182)	0.631 (0.119)	-0.029 (0.956)
Economics & Business	0.088 (0.789)	0.053 (0.897)	-0.174 (0.647)	0.071 (0.816)	-0.266 (0.574)	0.755* (0.091)	0.143 (0.737)
Birth order	-0.196 (0.115)	-0.190 (0.156)	-0.006 (0.970)	-0.164 (0.204)	0.157 (0.348)	-0.106 (0.499)	-0.214 (0.226)
Female	0.111 (0.713)	0.440 (0.134)	-0.075 (0.822)	0.100 (0.705)	0.278 (0.420)	0.878** (0.013)	0.498 (0.197)
Black	0.282 (0.321)	0.278 (0.306)	0.161 (0.596)	0.249 (0.365)	0.275 (0.434)	-0.292 (0.411)	0.300 (0.454)
Older than 21	-0.049 (0.862)	0.304 (0.267)	-0.585* (0.081)	-0.263 (0.307)	0.211 (0.510)	-0.002 (0.995)	-1.796*** (0.001)
Constant	-0.840* (0.084)	-0.389 (0.395)	0.208 (0.652)	-0.207 (0.627)	-0.697 (0.224)	-1.335** (0.031)	1.705** (0.047)
Observations Nr of subjects	116 116	80 40	80 40	80 40	76 38	78 39	76 38
Log-likelihood	-57.40	-51.89	-51.74	-48.54	-43.67	-41.13	-42.79
Within subjects							
(%) “Violations“ Observations {SR,RS}		52.50 {10, 11}	37.50 {9, 6}	50.00 {10, 10}	44.74 {14, 3}	35.90 {5, 9}	31.58 {9, 3}
Cochran’s chi2 (1) (Pr. > chi2) {Exact *p*}		0.048 (0.827) {1.000}	0.600 (0.439) {0.607}	0.00 (1.000) {1.000}	7.12*** (0.008) {0.013}	1.14 (0.285) {0.424}	3.00* (0.083) {0.146}

p-values in parentheses: *** if p<0.01; ** if p<0.05; * if p<0.1

Table A.2. Probit Tests of Hypothesis 2 (CCE)

Variables/Mechanism	**OT**	**PORnp**	**PORpi**	**PAC/N**	**PAC**	**PAS**	**PAI**
Pair 4 (D)	0.079 (0.790)	-0.422 (0.239)	-0.592* (0.058)	-0.446* (0.074)	-0.087 (0.800)	-1.082*** (0.002)	-0.060 (0.845)
Science & Engineering	-0.386 (0.297)	0.657*** (0.005)	0.343 (0.428)	0.290 (0.504)	0.082 (0.823)	0.748 (0.130)	-0.478 (0.327)
Economics & Business	-0.214 (0.527)	0.467 (0.250)	0.022 (0.950)	-0.315 (0.514)	-0.448 (0.319)	0.745 (0.157)	0.015 (0.973)
Birth Order	-0.318** (0.019)	-0.211* (0.093)	-0.179 (0.274)	0.051 (0.772)	-0.020 (0.901)	0.023 (0.907)	-0.375** (0.044)
Female	0.217 (0.479)	0.647** (0.013)	-0.030 (0.931)	0.592 (0.105)	-0.040 (0.906)	1.629*** (0.000)	0.645* (0.060)
Black	-0.058 (0.845)	0.378 (0.140)	0.177 (0.577)	0.303 (0.388)	0.655** (0.043)	-0.195 (0.628)	-0.010 (0.975)
Older than 21	-0.282 (0.327)	-0.013 (0.966)	-0.457 (0.127)	-0.689* (0.086)	-0.235 (0.461)	0.051 (0.899)	-0.536 (0.162)
Constant	0.263 (0.585)	-0.598* (0.096)	0.248 (0.665)	-0.527 (0.336)	-0.737 (0.164)	-1.906*** (0.005)	0.706 (0.296)
Observations Nr of subjects	96 96	80 40	80 40	80 40	76 38	78 39	76 38
Log-Likelihood	-53.07	-47.42	-46.37	-41.43	-37.26	-29.37	-43.87
Within subjects							
(%) “Violations“ Observations {SR,RS}		60.00 {15, 9}	45.00 {13, 5}	22.50 {7, 2}	34.21 {7, 6}	28.21 {10, 1}	39.47 {8, 7}
Cochran’s chi2 (1) (Pr. > chi2) {Exact *p*}		1.50 (0.221) {0.308}	3.56* (0.059) {0.096}	2.78 (0.096) {0.180}	0.08 (0.782) {1.000}	7.36*** (0.007) {0.012}	0.067 (0.796) {1.000}

p-values in parentheses: *** denotes p<0.01; ** denotes p<0.05; * denotes p<0.1

Table A.3. Probit Tests of Hypothesis 3 (DCRE)

Variables / Mechanism	OT	PORnp	PORpi	PAC/N	PAC	PAS	PAI
Pair 3 (D)	-0.343 (0.215)	0.354 (0.213)	0.503** (0.038)	-0.082 (0.737)	-0.427** (0.047)	0.272 (0.358)	-0.025 (0.923)
Science & Engineering	-0.015 (0.965)	0.137 (0.714)	-0.174 (0.737)	0.830** (0.045)	0.298 (0.483)	0.332 (0.492)	-0.310 (0.554)
Economics & Business	-0.138 (0.673)	-0.894*** (0.009)	-0.479 (0.292)	0.834* (0.053)	-0.597 (0.325)	0.520 (0.264)	-0.330 (0.484)
Birth Order	-0.171 (0.186)	-0.265 (0.187)	-0.103 (0.598)	-0.160 (0.336)	0.122 (0.553)	0.153 (0.377)	-0.439** (0.044)
Female	-0.010 (0.971)	1.372*** (0.000)	0.203 (0.582)	0.448 (0.217)	-0.367 (0.366)	1.328*** (0.001)	0.289 (0.436)
Black	0.134 (0.626)	-0.122 (0.713)	0.542 (0.167)	0.453 (0.206)	0.553 (0.167)	-0.787* (0.076)	0.296 (0.460)
Older than 21	-0.075 (0.787)	-0.102 (0.805)	-0.866** (0.027)	-0.858** (0.023)	-0.196 (0.633)	-0.094 (0.806)	-0.969* (0.057)
Constant	0.134 (0.780)	-0.566 (0.303)	-0.185 (0.776)	-0.382 (0.419)	-0.482 (0.440)	-1.761*** (0.002)	1.354 (0.109)
Observations Nr. of subjects	96 96	80 40	80 40	80 40	76 38	78 39	76 38
Log-likelihood	-58.63	-42.09	-43.75	-43.02	-40.94	-38.64	-42.98
Within subjects							
(%) “Violations“ Observations {SR,RS}		30.00 {4, 8}	25.00 {2, 8}	22.50 {5, 4}	18.42 {6, 1}	28.21 {4, 7}	26.32 {5, 5}
Cochran’s chi2 (1) (Pr. > chi2) {Exact *p*}		1.33 (0.248) {0.388}	3.6* (0.058) {0.109}	0.11 (0.738) {1.000}	3.57* (0.059) {0.125}	0.818 (0.366) {0.549}	0.00 (1.000) {1.000}

p-values in parentheses: *** denotes p<0.01; ** denotes p<0.05; * denotes p<0.1

Table A.4. Probit Tests of Hypothesis 4 (DCCE)

Variables / Mechanism	**OT**	**PORnp**	**PORpi**	**PAC/N**	**PAC**	**PAS**	**PAI**
Pair 5 (D)	0.707** (0.023)	0.394 (0.109)	0.005 (0.980)	0.291 (0.278)	-0.293 (0.268)	-0.168 (0.519)	-0.007 (0.979)
Science & Engineer	0.618 (0.103)	-0.204 (0.564)	-0.131 (0.773)	0.380 (0.374)	0.437 (0.269)	0.232 (0.614)	0.376 (0.505)
Economics & Business	0.507 (0.183)	-0.011 (0.985)	-0.451 (0.293)	-0.126 (0.739)	-0.105 (0.825)	1.098** (0.021)	0.607 (0.162)
Birth Order	-0.012 (0.933)	0.130 (0.435)	0.104 (0.581)	-0.231 (0.195)	0.003 (0.987)	-0.361 (0.104)	0.056 (0.782)
Female	-0.165 (0.647)	0.192 (0.580)	-0.013 (0.974)	0.469 (0.157)	0.514 (0.140)	0.257 (0.521)	-0.304 (0.438)
Black	0.624* (0.075)	0.255 (0.446)	-0.027 (0.943)	0.018 (0.958)	-0.467 (0.197)	-0.138 (0.756)	0.449 (0.260)
Older than 21	0.100 (0.757)	0.542 (0.126)	-0.151 (0.667)	0.338 (0.369)	-0.079 (0.805)	0.332 (0.424)	…
Constant	-1.689*** (0.002)	-0.715 (0.185)	0.093 (0.896)	-0.429 (0.438)	-0.052 (0.929)	-0.718 (0.325)	-0.712 (0.353)
Observations Nr. of subjects	97 97	80 40	80 40	80 40	76 38	78 38	66 33
Log-likelihood	-47.24	-52.27	-54.04	-48.58	-49.05	-34.58	-42.61
Within subjects							
(%) “Violations“ Observations {SR,RS}		35.00 [4, 10}	25.00 {5, 5}	35.00 {5, 9}	36.84 {9, 5}	15.38 {4, 2}	26.32 {5, 5}
Cochran’s chi2 (1) (Pr. > chi2) {Exact *p*}		2.57 (0.101) {0.180}	0.00 (1.000) {1.000}	1.14 (0.285) {0.424}	1.14 (0.285) {0.424}	0.67 (0.414) {0.688}	0.00 (1.000) {1.000}

… the variable was dropped because it predicted choices perfectly

p-values in parentheses: *** denotes p<0.01; ** denotes p<0.05; * denotes p<0.1

Table A.5. Probit Regressions with PORpas Data

VARIABLES	Probit 3 (PORpas, PAS, PORpi & PORnp data)			EU Paradoxes (PORpas Data)		Cross-Task Effects (PORpas Data)
	All Rounds	Round 1	Round 5	CRE	CCE	Accumulated and Preceding Payoffs
Science & Engineering	0.093 (0.441)	0.077 (0.645)	0.003 (0.984)	0.216 (0.718)	-0.078 (0.877)	0.003 (0.994)
Economics & Business	0.088 (0.516)	0.205 (0.228)	0.138 (0.421)	-0.708 (0.204)	-0.219 (0.637)	-0.348 (0.374)
Birth Order	-0.071 (0.180)	-0.163** (0.016)	-0.125* (0.068)	0.890*** (0.001)	0.048 (0.831)	0.014 (0.927)
Female	0.425*** (0.000)	0.176 (0.226)	0.366** (0.014)	2.376*** (0.002)	…	0.771** (0.024)
Black	-0.025 (0.822)	0.175 (0.212)	0.070 (0.621)	-1.883*** (0.005)	-0.780* (0.069)	-0.619* (0.099)
Older than 21	0.089 (0.431)	0.147 (0.296)	-0.084 (0.556)	1.594** (0.013)	0.379 (0.445)	0.815*** (0.007)
DPORnp	0.519*** (0.000)	0.747*** (0.001)	0.574** (0.013)			
DPORpi	0.297** (0.044)	0.218 (0.336)	0.508** (0.022)			
DPAS	-0.151 (0.365)	-0.120 (0.622)	-0.210 (0.399)			
DPORpas	-0.073 (0.689)	0.101 (0.665)	-0.404 (0.118)			
EV Difference	-0.157 (0.213)	-0.055 (0.781)	-0.143 (0.467)			0.223 (0.442)
VAR Difference	0.029** (0.044)	0.009 (0.708)	0.010 (0.664)			-0.038 (0.284)
Pair 3 (D)				-0.181 (0.547)	-0.560 (0.139)	
Accumulated Payoff						-0.002 (0.806)
Preceding Payoff						-0.000 (0.991)
Constant	-0.901*** (0.000)	-0.635** (0.013)	-0.524** (0.044)	-4.829*** (0.000)	-0.027 (0.973)	-1.214** (0.024)
Observations Nr. of subjects	1,026 390	390 390	390 390	80 34	48 24	200 40
Log-likelihood	-611.3	-232.1	-226.4	-21.64	-27.49	-97.80

… the variable was dropped because it predicted choices perfectly
p-values in parentheses: *** denotes p<0.01; ** denotes p<0.05; * denotes p<0.1

Table A.6. Probit Regressions for Tasks 2, 3 and 4

VARIABLES	Probit 1	Probit 2	Probit 3	EU Paradoxes (Impure OT Paid Data)	
				CRE	CCE
EV Difference	-0.081 (0.399)		-0.096 (0.335)		
VAR Difference	0.028** (0.013)		0.031*** (0.007)		
Science & Engineering		0.254** (0.018)	0.257** (0.018)	-0.251 (0.576)	0.345 (0.439)
Economics & Business		0.128 (0.276)	0.133 (0.262)	-0.015 (0.975)	0.520 (0.324)
Birth Order		-0.112** (0.018)	-0.115** (0.017)	-0.411* (0.073)	-0.230 (0.311)
Female		0.427*** (0.000)	0.435*** (0.000)	0.018 (0.966)	0.540 (0.199)
Black		0.041 (0.669)	0.049 (0.613)	-0.336 (0.397)	-0.452 (0.293)
Older than 21		-0.051 (0.596)	-0.058 (0.551)	0.531 (0.205)	-0.136 (0.738)
DPORnp	0.538*** (0.001)	0.536*** (0.001)	0.558*** (0.001)		
DPORpi	0.449*** (0.009)	0.445** (0.010)	0.467*** (0.008)		
DPAC	0.289* (0.092)	0.357** (0.038)	0.380** (0.030)		
DPAC/N	0.244 (0.177)	0.329* (0.066)	0.353* (0.052)		
DPAS	-0.025 (0.899)	0.029 (0.878)	0.045 (0.817)		
DPAI	0.526*** (0.004)	0.590*** (0.001)	0.619*** (0.001)		
DPORpas	-0.035 (0.872)	-0.117 (0.580)	-0.095 (0.657)		
DimpureOT	0.362* (0.052)	0.395** (0.038)	0.417** (0.030)		
Pair 3 (D)				0.415 (0.282)	0.406 (0.287)
Constant	-0.906*** (0.000)	-0.902*** (0.000)	-1.098*** (0.000)	0.213 (0.762)	0.224 (0.753)
Observations Nr. of Subjects	1,056 506	1,056 506	1,056 506	52 52	51 51
Log-likelihood	-637.6	-623.5	-617.4	-30.42	-31.11

p-values in parentheses: *** denotes p<0.01; ** denotes p<0.05; * denotes p<0.1