European Economic Review 101 (2018) 20–34

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/euroecorev

Separated decisions☆

Alexander L. Brown[a], Paul J. Healy[b,*]

[a] Department of Economics, Texas A&M University, United States
[b] Department of Economics, The Ohio State University, United States

ARTICLE INFO

Article history:
Received 10 February 2017
Accepted 21 September 2017
Available online 5 October 2017

JEL classification:
C90
D01
D03
D81

Keywords:
Payment mechanism
Experimental methodology
Monotonicity
Decisions under uncertainty

ABSTRACT

We use experiments to test the incentive compatibility of the "random problem selection" payment mechanism, in which only one choice out of many is randomly chosen for payment. We find that the mechanism is not incentive compatible when all decisions are shown together in a single list. But when the rows of the list are randomized and shown on separate screens, incentive compatibility is restored. This causes more apparent intransitivities in choice ("multiple switching"), but, since the experiment is incentive compatible, these intransitivities must be inherent in subjects' preferences.

1. Introduction

Consider an experiment in which subjects make multiple decisions. If they are paid for every decision, payments from one decision may affect their preferred choice in another. It has been proposed that the random problem selection (RPS) mechanism avoids this problem by paying for only one randomly-chosen decision.[1] Recent theoretical work has achieved a fairly complete understanding of the conditions under which this mechanism is "incentive compatible", meaning observed choices accurately reflect underlying preferences.[2] But empirical tests of incentive compatibility for the RPS mechanism have generated mixed results. We provide here a reconciliation of these mixed results: When choices are shown together in a list, incentive compatibility is sometimes violated in the data. When choices are separated (for example, each displayed on a separate computer screen), incentive compatibility is restored.

☆ The authors thank Yaron Azrieli, Christopher P. Chambers, Jim Cox, Yoram Halevy, and Glenn Harrison for their helpful comments and conversations. We also thank Jim Cox, Vjollca Sadiraj, and Ulrich Schmidt for sharing their data. Healy gratefully acknowledges the NSF for funding this project under Award #SES-0847406.

* Corresponding author.
E-mail addresses: alexbrown@tamu.edu (A.L. Brown), healy.52@osu.edu (P.J. Healy).

[1] This mechanism has many names. Perhaps the most common is the "random lottery incentive mechanism" (Safra et al., 1990). We adopt RPS—which comes from Beattie and Loomes (1997)—because the Azrieli et al. (2016) framework we employ does not require randomness to be represented by objective lotteries.

[2] See the characterizations of Azrieli et al. (2016), and the references listed in footnote [12] below.

https://doi.org/10.1016/j.euroecorev.2017.09.014
0014-2921/

We generate these results using two different experiments. In the first we test the RPS mechanism in a multiple price list setting often used for eliciting risk preferences (e.g., Holt and Laury, 2002). Our multiple price list consists of 20 binary choices over lotteries, presented on a computer screen as one large list. Our experiment (whose design we describe below) shows that the RPS mechanism is not incentive compatible in this setting. Subjects do not reveal their underlying preferences truthfully.

We hypothesize that the failure of the RPS mechanism is due to the presentation of the problems in a list format. In our second experiment we replicate the first experiment exactly, except that we randomize the order of the 20 binary choices and show each on a separate screen. In this setting our data cannot reject incentive compatibility of the RPS mechanism. Our review of the literature (see Section 6) is consistent with this conclusion: the only tests that have shown the RPS mechanism to fail also use a list presentation.

In our experiments we use a test of incentive compatibility that controls for framing effects in a way that many other tests in the literature do not. To illustrate, consider the following example of an experimental test employing a “One Choice” treatment and an “RPS” treatment. Subjects in the One Choice treatment choose one option from the set $D_1 = \{x, y\}$. Subjects in the RPS treatment make choices from both $D_1 = \{x, y\}$ and $D_2 = \{x^+, y^-\}$, where x^+ is clearly better than x, and y^- is clearly worse than y. For example, x is an average-priced glass of Cabernet and x^+ is an expensive Cabernet (both offered for free). And y is an average-priced Merlot and y^- is a cheap Merlot (again, both free).[3] If, in D_1, subjects predominantly choose x (average Cabernet) in the One Choice treatment but choose y (average Merlot) in the RPS treatment, then one might be tempted to conclude that the “true” preference is for the Cabernet, and the RPS mechanism causes subjects to misreport that preference. In other words, it is not incentive compatible. But an alternative explanation is that the typical subject in the One Choice treatment prefers the Cabernet while the typical subject in the RPS treatment truly prefers the Merlot. This could arise because the presence of the expensive Cabernet makes the average Cabernet look less appealing while the presence of the cheap Merlot makes the average Merlot look more appealing. The “true” preferences over D_1 are altered by the presence of D_2. We refer to this as a *framing effect*. And note that an experiment like this cannot distinguish between an incentive compatibility failure and a framing effect.

To avoid this confound, the experimenter can add a third treatment in which subjects make choices in both D_1 and D_2, but are only paid for their choice from D_1. We refer to this as the “Framed Control” treatment. Paying only from D_1 ensures that subjects reveal their true favorite item in D_1. And the presence of D_2 presumably shifts preferences to $\succ'$ (instead of $\succ$), which is the same preference subjects have in the RPS treatment. Thus, if choices in the Framed Control treatment are different than in the RPS treatment, we know that it must be the payment mechanism that distorted choices.[4]

Similarly, comparing the Framed Control treatment to the One Choice treatment provides a clean test of the framing effect. Both treatments pay the same (only from D_1), but differ in whether subjects see D_2 or not. If subjects in the One Choice treatment reveal $x \succ y$ while subjects in the Framed Control treatment reveal $y \succ' x$, then we know that preferences are changed from $\succ$ to $\succ'$ by the addition of D_2.

In the experiments reported in this paper we run the three treatments described, but using a 20-row multiple price list. Thus, our actual experiment consists of twenty binary decisions $(D_1, \ldots, D_{20})$, not two. Our One Choice treatment has subjects view only the fourteenth row of the 20-row list (D_{14}) and make only that one choice. Our RPS treatment has subjects make choices from all 20 rows and selects one randomly for payment. Finally, our Framed Control treatment has subjects make choices from all 20 rows, but pays them only for their choice on the fourteenth row (D_{14}). Comparing choices from D_{14} across treatments provides our tests of incentive compatibility and framing.

In our first experiment we run all three treatments, presenting the twenty decisions as a single 20-row list on the computer screen. We find significant differences in D_{14} choices between the Framed Control treatment and the RPS treatment. This indicates a failure of incentive compatibility of the RPS mechanism; however, our p-Value for the test is 0.041, indicating a statistically significant but not overwhelming failure. In our second experiment we re-run the Framed Control and RPS treatments, but this time showing each choice on a separate screen and in random order. With this separated presentation we clearly fail to reject the null hypothesis of equality between treatments; our p-Value is 0.697.[5] Thus, the separated presentation restores incentive compatibility of the RPS mechanism.

We can also test for framing effects with and without the list presentation by comparing the Framed Control treatment to the One Choice treatment. Here we find marginal differences with the list presentation, and insignificant differences with the separated presentation. Thus, we conclude that there is suggestive evidence of a possible framing effect under the list

[3] This example is based on two of the treatments run by Cox et al. (2014a), where x, y, x^+, and y^- are lotteries, x^+ stochastically dominates x, and y^- is stochastically dominated by y.

[4] This assumes that the mechanism itself does not alter underlying preferences; we refer to this as “mechanism invariance”. Also, the Framed Control treatment is designed only for the purpose of testing incentive compatibility; it obviously wouldn’t be useful in other experiments where choices from D_2 are of interest.

[5] One concern is the power of our statistical test. We chose our sample size of 60 subjects per treatment *ex-ante*, based on a power calculation in which we targeted 70%% power and assumed the effect sizes found in Starmer and Sugden (1991). Specifically, we assumed the proportion of subjects choosing the riskier lottery in D_{14} in our Framed Control and RPS treatments would mirror those of Treatments C (55%%) and D (32.5%%), respectively, from that paper. We chose to use the Chi-squared test because it can be partitioned in a way that allows us to test all three treatments; details appear below. The choice of 70%% power (instead of 80%%) was driven by budget concerns.

presentation. But we stress that our sample size was chosen to test our primary hypotheses about incentive compatibility—not these framing hypotheses—so replications may be needed to verify or falsify our framing results.

In practice, most experimenters care about *both* incentive compatibility and framing effects, because they want to interpret each observed choice in the RPS mechanism as being identical to that same choice being made in isolation.[6] Thus, comparison of the One Choice treatment to the RPS treatment is of interest in practice. But that comparison alone does not tell us whether differences are due to the payment mechanism or framing effects. So it cannot be used to make recommendations about which payment mechanism is appropriate to use. Adding the Framed Control treatment allows us to disentangle the two effects and make clear recommendations about the RPS mechanism.

Beattie and Loomes (1997), Cubitt et al. (1998, Experiment 2), Freeman et al. (2012), Cox et al. (2014a), and Harrison and Swarthout (2014) all present experiments in which a One Choice treatment is compared to an RPS treatment, without the Framed Control. Thus, their tests are joint tests of incentive compatibility and framing effects.[7] Overall, results from these papers are mixed.

To our knowledge there are three previous experiments featuring a Framed Control treatment, run by Starmer and Sugden (1991), Cubitt et al. (1998, Experiment 3), and Cox et al. (2014b). We describe each in Section 6. Results are mixed: Starmer and Sugden (1991) find that the RPS mechanism is incentive compatible in one test, but not the other. The other two papers find no significant violation of incentive compatibility.[8] We organize these findings by noting that the Starmer and Sugden (1991) rejection occurs when choices are presented in a list, but the other two studies—which find no rejection—present choices in a separated format.

Why does the list presentation cause the RPS mechanism to fail? We conjecture that it induces subjects to treat the list of decisions as one large decision.[9] In doing so, subjects' choices become more consistent with the reduction of compound lotteries. It is well known that if a subject satisfies reduction but violates expected utility, then they must violate the axiom of monotonicity.[10] Azrieli et al. (2016) show that, theoretically, monotonicity is crucial for the RPS mechanism to be incentive compatible. So if any of our subjects have non-expected utility preferences but were induced to satisfy reduction because of the list presentation, then they would have generated the differences across treatments that we observed. The separated presentation may prevent reduction from being satisfied, in which case violations of expected utility have no consequence for the (theoretical) incentive compatibility of the RPS mechanism. Indeed, we find that, empirically, incentive compatibility is restored with the separated presentation.[11]

The idea of separating decisions has been proposed before, even in the context of multiple price lists. But previous authors dismissed its usage because the separated format generates greater inconsistency between choices (e.g., Eckel et al., 2005). Our conclusions turn this criticism on its head: The separated decisions give us an accurate view of subjects' underlying preferences, while the list format does not. Therefore we can conclude that underlying preferences in the separated format indeed have inconsistencies, but we cannot conclude that underlying preferences in the list presentation are consistent—even if the observed choices appear consistent—because incentive compatibility is violated.

In Section 3 we describe in detail our first experimental design with the list presentation. We carefully review all assumptions necessary to achieve a test of incentive compatibility and then show that, in our data, incentive compatibility fails. Then, in Section 4 we describe our second experiment with the separated presentation. Our results here show that incentive compatibility is restored. In Section 5 we identify behavioral differences between the list and separated formats, focusing mainly on inconsistencies in choices between adjacent rows of the list. First, we find that the list format generates more consistent *choices*, but without incentive compatibility we cannot be sure that underlying *preferences* are actually consistent. Second, the aggregate choice frequencies between the two formats do not significantly differ. Curiously, the incentive compatibility failures in the list format are almost entirely offset by a list framing effect, so that overall behavior appears identical between the list format and the separated format. Whether this conclusion would generalize is an open question, but seems unlikely as it requires an exact offsetting of two seemingly unrelated phenomena.

In the last section we compare our results to the previous tests of incentive compatibility described above, restricting attention to those that feature a Framed Control treatment. We see that the only rejections of incentive compatibility occur when decisions are presented in a list format. Thus, the previous literature corroborates our empirical finding that the RPS mechanism may not be incentive compatible when decisions are presented as a list, but appears to be incentive compatible when decisions are separated.

[6] This is not always true; in some cases experimenters are explicitly interested in framing effects.

[7] Cox et al. (2014a) also run an RPS treatment in which subjects choose (1) between x and y, and (2) between x^- and y^+, where x^- is dominated by x and y^+ dominates y. Choice frequencies of x and y are similar to the treatment in which subjects only see x and y. It may be that x^- and y^+ do not cause any framing effect and the RPS mechanism is incentive compatible. Or it may be that there is a framing effect, but it is 'offset' by misreports due to the RPS mechanism. In our own experiment we find a similar pattern of offsetting effects.

[8] Cox et al. (2014b) avoid the framing confound when comparing their "ImpureOT" to "POR" treatments, but not when comparing "OT" to "POR." They do find significant differences across various mechanisms, but we focus here only on the ImpureOT vs. POR comparison of interest.

[9] This conjecture also appears in Freeman et al. (2012).

[10] A formal definition of monotonicity appears below.

[11] We reiterate that this is simply a conjecture consistent with the data. Our design does not provide a test for either reduction or non-expected utility preferences.

2. Theoretical background

We adapt the theoretical framework of Azrieli et al. (2016) to study incentives in our experiment.[12] A subject is given twenty decision problems, denoted D_1 through D_{20}. Each D_i offers a choice between a "safe" lottery l_0 that is the same in every problem, and a "risky" lottery l_i that varies across problems. Thus, $D_i = \{l_0, l_i\}$ for each i. Let $L = \{l_0, l_1, \ldots, l_{20}\}$ be the set of all lotteries appearing in the experiment. We assume the subject has a complete, reflexive, and transitive preference relation $\succeq$ over L that represents the choices they would make from any subset of L. We, as experimenters, want to learn $\succeq$. Specifically, we want to know, for each D_i, whether $l_0 \succeq l_i$ or $l_i \succeq l_0$.[13]

The lotteries in L are payments that depend on the draw from a bingo cage containing 20 balls.[14] Letting $B = \{1, \ldots, 20\}$ denote the set of balls in the cage, each lottery l is a function that pays $l(b)$ dollars when ball $b \in B$ is drawn. We want l_0 to be an equiprobable gamble between \$10 and \$5, so we set

$$l_0(b) = \begin{cases} \$10 & \text{if} \quad b \leq 10 \\ \$5 & \text{if} \quad b > 10 \end{cases}$$

For each risky lottery l_i (where $i \in \{1, \ldots, 20\}$), we set

$$l_i(b) = \begin{cases} \$15 & \text{if} \quad b \leq i \\ \$0 & \text{if} \quad b > i \end{cases}$$

If $\succeq$ respects dominance then $l_i \succ l_j$ if and only if $i > j$, meaning there must be a unique switch point i^* such that $l_0 \succ l_i$ for $i < i^*$, $l_i \succeq l_0$ for $i = i^*$, and $l_i \succ l_0$ for $i > i^*$.[15] In other words, when working through the problems sequentially, there will be a unique problem i^* at which the subject switches from choosing l_0 to choosing l_i. If the subject is a risk-neutral expected utility maximizer, then $i^* = 10$. In general, we do not assume that $\succeq$ satisfies expected utility, or even dominance. In fact, $\succeq$ may not even satisfy probabilistic sophistication (Machina and Schmeidler, 1992). For this reason we work in a purely subjective framework (using states, as in Savage, 1954), rather than an objective lotteries framework (using probabilities, as in Von Neumann and Morgenstern, 1944).

Let $c_i \in D_i$ represent the subjects' choice in each D_i. For the RPS mechanism, we roll a twenty-sided die to determine which c_i is paid out. When the die roll is $\omega \in \Omega = \{1, \ldots, 20\}$, the subject receives lottery $c_\omega \in D_\omega \subset L$. If the die roll is ω and the bingo ball drawn in b, then the final dollar payment received is $c_\omega(b)$. Thus, the subject's twenty choices constitute a two-stage act. For each vector of choices $c = (c_1, \ldots, c_{20})$, let $\langle c_1, \ldots, c_{20} \rangle$ denote that two-stage act, where the ith entry specifies the (single-stage) lottery paid if row i is chosen by the 20-sided die for payment. The space of two-stage acts is given by L^Ω.

To describe which vector of choices c the subject will pick, we must describe her preferences over two-stage acts. Thus, we need to "extend" $\succeq$ over single-stage acts (L) to the space of two-stage acts (L^Ω). Let $\trianglerighteq$ denote this extension of $\succeq$, with $\triangleright$ denoting the strict relation. We refer to $\trianglerighteq$ as the subject's "two-stage" preferences, and $\succeq$ as their "second-stage" preferences.

Recall that the experimenter is interested in learning $\succeq$ over L, but subjects actually make choices over two-stage acts according to $\trianglerighteq$. Thus, we want the subject's choice over two-stage acts (given by $\trianglerighteq$) to inform us about her underlying preference over L (given by $\succeq$). More concretely, we want her choice vector c (chosen according to $\trianglerighteq$) to reveal her true favorite element of each D_i (based on $\succeq$). We say an announcement c^* is *truthful for* $\succeq$ if, for each i and $c'_i \in D_i$, we have $c^*_i \succeq c'_i$. The RPS mechanism is *incentive compatible* if the truthful message generates the most-preferred two-stage act.[16]

Definition 1 (Incentive compatibility of the RPS mechanism)**.** The RPS mechanism is *incentive compatible* if, for any preference $\succeq$, any $c^* \in \times_i D_i$ that is truthful for $\succeq$ and any $c' \in \times_i D_i$, we have $\langle c^*_1, \ldots, c^*_{20} \rangle \trianglerighteq \langle c'_1, \ldots, c'_{20} \rangle$. Additionally, if c' is not truthful for $\succeq$ then $\langle c^*_1, \ldots, c^*_{20} \rangle \triangleright \langle c'_1, \ldots, c'_{20} \rangle$.

We require strict incentive compatibility: truth-telling is optimal and strictly preferred to any non-truthful announcement.

[12] Azrieli et al. (2016) follows a long line of research on this topic. Important contributions include Holt (1986), Karni and Safra (1987), Segal (1988, 1990), Oechssler and Roomets (2014), Oechssler et al. (2016), Baillon et al. (2014), Bade (2012), Kuzmics (2017), and Azrieli et al. (2012), among others. Discussions of the RPS mechanism date back to the 1950s, in the works and discussions of Wold, Savage, Allais, and Wallis. Early applications include Becker et al. (1964), Yaari (1965), and Grether and Plott (1979).

[13] The preference relation only applies to the current experiment with these twenty problems presented in a specific way. If the decision problems or their presentation were changed, the preference relation could change as well. This is what we call a framing effect.

[14] Bingo is a game of chance popular in the United States. Random numbers in Bingo are typically generated by drawing numbered balls from a rotating spherical metal cage, called a "bingo cage". We use an actual bingo cage in our experiment.

[15] We assume all subjects prefer more money to less. Dominance then says that if $l(b) \geq l'(b)$ for every b and there is some b' such that $l(b') > l'(b')$ then $l \succ l'$.

[16] For expositional simplicity, our definition of incentive compatibility is specific to the RPS mechanism. See Azrieli et al. (2016) for a generalized definition that applies to any payment mechanism.

Table 1
In the RPS mechanism, a truthful choice c^* dominates any non-truthful choice c'. Here, c' gives less-preferred lotteries for rolls 14 and 15. Monotonicity therefore ensures that c^* is chosen over c'.

Die roll (ω)	1	$\cdots$	13	14	15	16	$\cdots$	20
Truth (c^*)	l_0	$\cdots$	l_0	l_{14}	l_{15}	l_{16}	$\cdots$	l_{20}
Non-truth (c')	l_0	$\cdots$	l_0	l_0	l_0	l_{16}	$\cdots$	l_{20}

2.1. Consistency and monotonicity

If we do not assume any link between $\trianglerighteq$ and $\succeq$, then choices in the experiment may not have any relationship to preferences over L. In that case incentive compatibility cannot be assured. The study of incentive compatibility of payment mechanisms is therefore a study of the possible links between $\trianglerighteq$ and $\succeq$. And any test of incentive compatibility is a test of these links.

The most basic assumption about the link between $\trianglerighteq$ and $\succeq$ is *consistency*. This means $\trianglerighteq$ agrees with $\succeq$ when we look at the subset of "degenerate" two-stage acts that pay the same second-stage lottery for every die roll ω.

Definition 2 (Consistency)**.** A subject satisfies *consistency* if, for any second-stage lotteries l_i and l_j, $l_i \succeq l_j$ if and only if $\langle l_i, \ldots, l_i\rangle \trianglerighteq \langle l_j, \ldots, l_j\rangle$ (equivalently written as $\langle l_i\rangle \trianglerighteq \langle l_j\rangle$).

We assume consistency throughout. Azrieli et al. (2016) show that consistency alone does not guarantee incentive compatibility of any experiment in which subjects make more than one choice.

A stronger link that does guarantee incentive compatibility is *monotonicity*.

Definition 3 (Monotonicity)**.** A subject satisfies *monotonicity* if $c_i \succeq c_i'$ for every i implies $\langle c_1, \ldots, c_{20}\rangle \trianglerighteq \langle c_1', \ldots, c_{20}'\rangle$, and if there is some i for which $c_i \succ c_i'$ then $\langle c_1, \ldots, c_{20}\rangle \triangleright \langle c_1', \ldots, c_{20}'\rangle$.

Monotonicity simply says that if choice vector c gives something better than c' for every roll of the die, then announcing c is preferred to announcing c'. It is equivalent to the compound independence axiom studied by Segal (1990) and others, but should not be confused with the mixture independence axiom of expected utility. The mixture independence axiom says that if $l \succeq l'$ then $p \cdot l + (1-p) \cdot l'' \succeq p \cdot l' + (1-p) \cdot l''$, where $p \cdot l + (1-p) \cdot l''$ represents a simple lottery in L that is a convex combination of l and l''. The mixture independence axiom only applies to $\succeq$. It places no restrictions on $\trianglerighteq$, and therefore has no immediate consequences for incentive compatibility.

Azrieli et al. (2016) show that the RPS mechanism is essentially the only incentive compatible mechanism if and only if $\trianglerighteq$ satisfies monotonicity.[17] To see the sufficiency of monotonicity, consider the example in Table 1. Suppose the truthful announcement c^* (shown in the top row) has the subject "switching" from the safe lottery l_0 to the risky lottery l_i at $i = 14$. Compare that to a non-truthful announcement c' (the bottom row) in which the subject instead switches at $i = 16$. Announcing c' gives the subject the same second-stage lotteries as c^* for die rolls 1 through 13 and die rolls 16 through 20, but gives strictly less-preferred lotteries for rolls 14 and 15. Thus, c' is dominated by c^*. Under monotonicity c' would never be chosen, because it is dominated. Since all non-truthful announcements are similarly dominated by the truthful announcement, the RPS mechanism is incentive compatible whenever monotonicity is assumed.[18]

Monotonicity appears to be a weak assumption. But it becomes strong when paired with another axiom that further restricts the relationship between $\trianglerighteq$ and $\succeq$, such as the reduction of compound lotteries. In this definition, recall that $\langle l_1, \ldots, l_{20}\rangle$ is a two-stage act, while $\Sigma_i p_i \cdot l_i$ is a convex combination of single-stage acts.

Definition 4 (Reduction)**.** A subject satisfies *reduction* if there is a probability distribution $p = (p_1, \ldots, p_{20})$ over Ω such that, for any $c = (l_1, \ldots, l_{20})$ and $c' = (l_1', \ldots, l_{20}')$,

$$\langle l_1, \ldots, l_{20}\rangle \trianglerighteq \langle l_1', \ldots, l_{20}'\rangle \iff \sum_{i=1}^{20} p_i \cdot l_i \succeq \sum_{i=1}^{20} p_i \cdot l_i'.$$

To see the strength of reduction when combined with monotonicity, imagine an experiment with $D_1 = \{l, l'\}$ and $D_2 = \{l'', l'''\}$ that uses the RPS mechanism. Suppose the truthful announcement has $c_1^* = l$, meaning $l \succeq l'$. Monotonicity implies that $\langle l, l''\rangle \trianglerighteq \langle l', l''\rangle$. Reduction then implies that $p \cdot l + (1-p) \cdot l'' \succeq p \cdot l' + (1-p) \cdot l''$. We've just shown that $l \succeq l'$ implies $p \cdot l + (1-p) \cdot l'' \succeq p \cdot l' + (1-p) \cdot l''$, so $\succeq$ in fact satisfies the mixture independence axiom of expected utility. This conclusion holds generally:

[17] If the experiment has only one decision problem then the RPS mechanism simply requires that the chosen option be paid with certainty, and this is clearly incentive compatible. We say "essentially" because, in theory, there can be other RPS-like mechanisms that are incentive compatible when the experiment generates "surely-identified sets" (Azrieli et al., 2016). In our experiment (and in most experiments) surely identified sets do not exist, so these mechanisms are not available. Thus, the RPS mechanism is the only one that is incentive compatible assuming monotonicity.

[18] The assumption of monotonicity also rules out the possibility that the subject perceives correlation between which lottery is drawn, and the outcome of any lotteries. If subjects perceive such correlation, then the RPS mechanism may not be incentive compatible.

Observation 1 (Segal, 1990)**.** If a subject satisfies both monotonicity and reduction, then their preference $\succeq$ over second-stage lotteries satisfies the independence axiom of expected utility.

We know that people exhibit violations of expected utility in certain settings. If in some of these settings they satisfy reduction, then they must violate monotonicity. Since monotonicity is necessary for incentive compatibility of the RPS mechanism, the mechanism will not be incentive compatible in those settings.[19]

2.2. Framing

We follow a broad, commonly-used definition of framing, popularized by Tversky and Kahneman (1981). As the authors note, one's "conception" of a decision problem (i.e., the "frame") may vary over the same decision problem due to "the formulation of the problem." When this differential framing leads to a change in choices (thus, a change in preferences), we have a "framing effect."[20]

Formally, we say a *framing effect* exists between two experiments if $\succeq$ differs between them in any observable way. For example, if $l_{14} \succ l_0$ in one experiment but $l_0 \succ l_{14}$ in another, then we say a framing effect has occurred. Again, we remain agnostic as to the cause of the framing effect and make no predictions about when they will or will not occur; we simply label any change in underlying preferences as a framing effect.

Importantly, framing effects do not represent a violation of incentive compatibility. If a framing effect exists between two experiments but both are incentive compatible then the change in preferences will be observed correctly. For this study we are particularly interested in a possible framing effect generated by showing the decision problems in a list.

Definition 5 (List framing effect)**.** A subject exhibits a *list framing effect* if $\succeq$ differs between an experiment in which only one D_i is given and an experiment in which $(D_1, \ldots, D_{20})$ are all displayed as an ordered list of problems.

Recall from the introduction that comparing an RPS treatment to a One Choice treatment confounds incentive compatibility failures with framing effects. And this confound can be avoided by adding a Framed Control treatment in which subjects face D_1 through D_{20} but are paid only for D_{14}. Comparing the Framed Control to the RPS treatment gives a clean test of incentive compatibility. But suppose that the payment mechanism itself alters preferences. Since the Framed Control and RPS treatments use different payment mechanisms, they could induce different underlying preferences. Even if the RPS mechanism were incentive compatible, the experimenter could still observe different choices across these treatments. Thus, for this comparison to be a clean test of incentive compatibility we need to rule out the possibility that the payment mechanism itself alters preferences. We refer to this as mechanism invariance.

Assumption 1 (Mechanism invariance)**.** A subject satisfies *mechanism invariance* if $\succeq$ does not differ between two experiments that are identical except for their payment mechanisms.

As mentioned in the introduction, we are not aware of any way to test mechanism invariance without first assuming incentive compatibility—and our test of incentive compatibility requires mechanism invariance—so all of our results should be read as being true conditional on the assumption of mechanism invariance.

3. Experiment 1: list format

3.1. Design

We recruited 181 subjects via email from the standard experimental economics subject pool at Ohio State University.[21] When a subject arrived in the lab, they were greeted by an experimenter and seated at a computer terminal. They signed a consent form and then received a printed questionnaire of Big 5 personality measures and other demographic questions. The experimenter emailed the subject a blank spreadsheet into which they typed their questionnaire answers. The subject then emailed the completed spreadsheet back to the experimenter. At that point, they were instructed to open a website on their computer that contains the decision-making interface.

When logging in to the website, subjects were randomly assigned to one of three treatments, summarized in Table 2. The treatment names are L-RPS (list display, RPS mechanism), L-14 (list display, pay only row 14), and O-14 (only row 14 shown, pay only row 14). These are our RPS, Framed Control, and One Choice treatments, respectively. Subjects randomly assigned

[19] Technically, monotonicity is only sufficient; a slightly weaker condition called ϕ-monotonicity is necessary. See Azrieli et al. (2016) for details.

[20] Additional research has become more specific on what types of framing effects are commonly found and how they interact with preferences, especially in the domain of risk (see, for example, Kahneman and Tversky, 1984 or Kahneman, 1992). However, we prefer the more broad definition of framing because it can capture any possible way in which preferences are altered due to differences in frames.

[21] This pool contains all Ohio State students who have recently enrolled in any Economics course, regardless of their major. It also includes anyone who voluntarily added themselves to the database, and excludes anyone who voluntarily removed themselves from the database. The dates of (and number of subjects in) each session were 10/21/2013 (39), 10/22/13 (39), 10/23/2013 (25), 2/26/2014 (44), and 3/3/2014 (34).

Table 2
The three treatments in Experiment 1. The results of L-RPS and L-14 should be equivalent under the assumption of monotonicity. The results of O-14 and L-14 should be equivalent if there is no list framing effect.

Treatment	L-RPS	L-14	O-14
Decisions (rows) shown	All 20	All 20	Only row 14
Decisions (rows) paid	One random	Only row 14	Only row 14
Display format	List	List	Single row

Row #	Option A			or	Option B		
1	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	☐	or	Ball 1 pays $15 (5% chance of $15)	Balls 2-20 pay $0 (95% chance of $0)	☐
2	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	☐	or	Balls 1-2 pay $15 (10% chance of $15)	Balls 3-20 pay $0 (90% chance of $0)	☐
3	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	☐	or	Balls 1-3 pay $15 (15% chance of $15)	Balls 4-20 pay $0 (85% chance of $0)	☐
4	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	☐	or	Balls 1-4 pay $15 (20% chance of $15)	Balls 5-20 pay $0 (80% chance of $0)	☐
	Balls 1-10 pay $10	Balls 11-20 pay $5			Balls 1-5 pay $15	Balls 6-20 pay $0	

Fig. 1. The display of the 20 decision problems in list format for treatments L-RPS and L-14 (cropped for space).

to treatment L-RPS saw the following instructions at the top of the website, followed by the table of decision problems shown in Fig. 1:

> **Please read the instructions before continuing**
> In the table below, each row presents two options: Option A and Option B. In every row, place a checkmark in the box to the right of the option you prefer to receive. Only one of your choices will be selected for payment. Specifically, the experimenter will roll a 20-sided die to determine which row will actually be used for payment. Then, s/he will play out your chosen option for that row by drawing one ball from a Bingo cage containing 20 balls. The number on the ball (1 through 20) will determine your actual payment for the experiment.
> For example, if, in the chosen row, the option you checked says "Balls 1–10 pay $10. Balls 11–20 pay $5", and if ball #17 is drawn, then you would be paid $5 for this experiment. It is possible to earn nothing for this experiment. If you have questions, please raise your hand.
> You must work through the rows in order. The software won't let you make a choice in one row until you've finished with the previous row. Also, you cannot submit your choices until all rows are complete.
> This is not a test, and there are no right or wrong answers. For each row, please choose the option that you prefer.
> Click to continue: [OK, I Read the Instructions]

Immediately below the instructions was the table of decision problems, shown (in part) in Fig. 1. The subject could not make any choices until they clicked the button labeled "OK, I Read the Instructions", and could not make a choice in a given row until they had made a choice in the previous row. They could make any choice in any row, and could switch back and forth between Option A and Option B as often as they wished. They were given no other instructions or advice on how to complete this task. At the bottom of the screen was a button labeled "Click Here When Finished," which was initially deactivated. Once the subject had made all 20 choices, the button became active and its label switched to "Submit Your Choices When Ready." Upon clicking this, their choices were recorded on the server and they saw a screen that read "Thank You. Your data has been recorded. If you have finished your questionnaire, then the experiment is finished. Please close your browser now and see the experimenter to determine your actual payment." At that point they went to the experimenter, who could now see their submitted decisions on his own computer screen. The experimenter rolled a 20-sided die to determine which row to pay, and drew a ball from a Bingo cage of 20 balls to determine the payout for the subject's chosen option in that row. The subject then received this payment in cash, plus a $5 show-up fee, filled out a payment receipt, and left the experiment.

Treatment L-14 (our Framed Control) is identical to L-RPS, except subjects were only paid for row 14. They still saw all 20 rows and were forced to make choices in each row, but they knew only the fourteenth row would be paid. Specifically, the first paragraph of the on-screen instructions read:

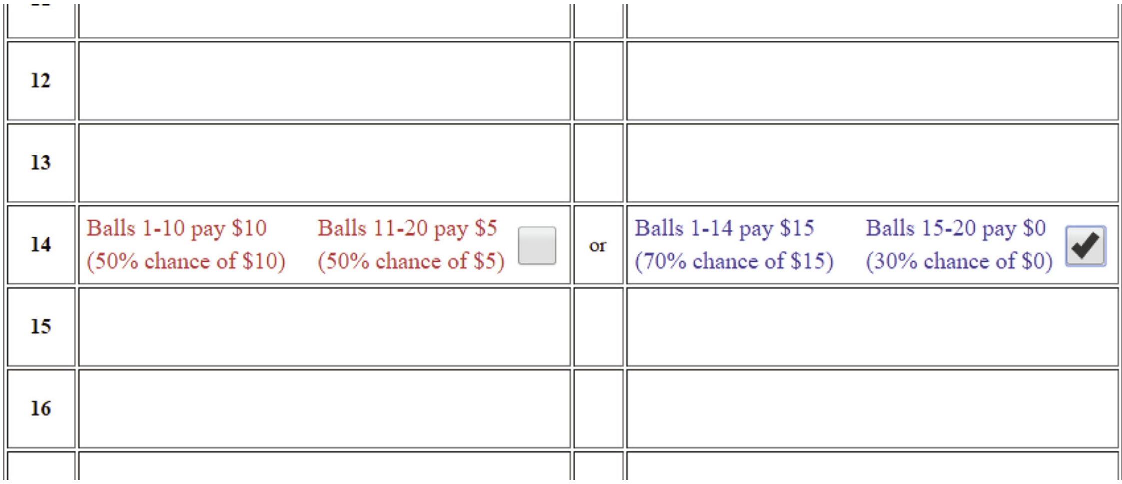

Fig. 2. The display of the one decision problem in O-14 (after a choice has been made).

> In the table below, each row presents two options: Option A and Option B. In every row, place a checkmark in the box to the right of the option you prefer to receive. You will be paid based on your choice in row 14. Specifically, the experimenter will play out your chosen option for that row by drawing one ball from a Bingo cage containing 20 balls. The number on the ball (1 through 20) will determine your actual payment for the experiment.

Every other aspect of the interface was exactly the same as in L-RPS, described above.[22] When the subject went to the experimenter for payment, the experimenter did not roll the 20-sided die since the payment row was pre-determined, but did use the Bingo cage to determine the subject's payout from their row-14 choice.

Finally, subjects in treatment O-14 (our One Choice treatment) saw a table of 20 rows, numbered from 1 to 20 as in the other treatments, but every row except row 14 was blank. The instructions were as follows, and a snippet of the choice interface is shown in Fig. 2.[23] All else was the same as L-14.

> **Please read the instructions before continuing**
>
> In the table below, row 14 presents two options: Option A and Option B. In that row, place a checkmark in the box to the right of the option you prefer to receive. The experimenter will play out your chosen option for that row by drawing one ball from a Bingo cage containing 20 balls. The number on the ball (1 through 20) will determine your actual payment for the experiment.
>
> For example, if the option you checked says "Balls 1–10 pay $10. Balls 11–20 pay $5", and if ball #17 is drawn, then you would be paid $5 for this experiment. It is possible to earn nothing for this experiment. If you have questions, please raise your hand.
>
> This is not a test, and there are no right or wrong answers. For each row, please choose the option that you prefer.
>
> Click to continue: [OK, I Read the Instructions]

3.2. Identifying assumptions

To lay out our identifying assumptions we begin with treatment O-14 (refer to Table 2 for a quick description of treatments). There, we observe the typical subject choosing the safe lottery l_0 over the risky lottery l_{14}. Technically, they are choosing the degenerate two-stage act $\langle l_0 \rangle$ over the degenerate two-stage act $\langle l_{14} \rangle$. Under the assumption of consistency, we can infer that $l_0 \succsim l_{14}$ in O-14.[24]

Suppose in L-14 we observe the opposite pattern of choice: $\langle l_{14} \rangle \trianglerighteq \langle l_0 \rangle$. Assuming consistency, we infer that $l_{14} \succsim l_0$. Because $\succsim$ differs between O-14 and L-14, we conclude that there is a list framing effect.

Now, if we additionally assume mechanism invariance, then $l_{14} \succsim l_0$ must also be true in L-RPS as well. Under monotonicity we should see subjects choosing l_{14} on the fourteenth row. But we don't; instead, the typical subject chooses l_0 over l_{14}. Thus, monotonicity is violated and the L-RPS mechanism is not incentive compatible.

To summarize, our test of the list framing effect assumes consistency, while our test of monotonicity assumes both consistency and mechanism invariance.

[22] The first sentence of the second paragraph was also changed to read "for example, if, in row 14, the option you checked says...", instead of "for example, if, in the chosen row, the option you checked says...".

[23] The figure shows the screen after the subject has chosen Option B; initially both checkboxes were blank.

[24] One might argue that we actually observe $l_0 \succsim l_{14}$, rather than $\langle l_0 \rangle \trianglerighteq \langle l_{14} \rangle$. If that is the interpretation then the consistency assumption is unnecessary.

Table 3
Percentage of subjects choosing the risky option (Option B) in row 14 of each treatment.

Treatment	L-RPS	L-14	O-14
%% Choosing risky	51.7%%	70.0%%	55.7%%
Sample size	60	60	61

Table 4
Logistic regressions with treatment L-14 omitted. *p*-Values in parentheses.

Variable	Reg.1	Reg.2	Reg.3
Constant	0.847***	1.022***	2.912
	(0.003)	(0.001)	(0.221)
List framing effect	−0.617	−0.650*	−0.716*
(Trt. O-14)	(0.106)	(0.091)	(0.071)
Monotonicity violation	−0.781**	−0.839**	−0.883**
(Trt. L-RPS)	(0.041)	(0.030)	(0.026)
Female	×	−0.387	−0.251
		(0.228)	(0.473)
Personality controls	×	×	✓
Either monotonicity violation or List framing effect (F-test of no treatment difference)	4.53 (0.1037)	5.05* (0.0801)	5.47* (0.0647)
Observations	181	181	180
Log likelihood	−120.08	−119.36	−116.88

*10%% significance, **5%% significance, ***1%% significance.

We could have tested O-14 against L-RPS directly (as was done by Freeman et al., 2012, Cox et al., 2014a, and others), but that would have required us to assume (incorrectly) that there is no list framing effect. Interestingly, we would have found no difference in that comparison—l_0 is chosen over l_{14} in both treatments—but that comparison would hide the fact that there is a list framing effect which happens to be offset almost entirely by the monotonicity violation in our data.

3.3. Results

The percentage of subjects choosing the risky lottery (Option B) in row 14 of each treatment is shown in Table 3. The riskier option is chosen more frequently in the L-14 treatment than the other two. A chi-squared test on the entire table shows that we cannot quite reject the hypothesis of identical frequencies across all three treatments (*p*-Value 0.0999), but if we partition the whole-table test into a test for monotonicity (L-RPS vs. L-14) and the remaining comparison (O-14 vs. {L-RPS $\cup$ L-14}), we find that the former difference is statistically significant (*p*-Value 0.0411) while the latter is insignificant (*p*-Value 0.0510) using the standard 0.05 threshold for significance.[25] Thus, we have evidence—though not overwhelming evidence—against the prior hypothesis of monotonicity.[26] Under our maintained assumption of mechanism invariance, we conclude that the RPS mechanism may not be incentive compatible in this experiment.

To validate these results, we run a logistic regression with the choice of the risky lottery in row 14 as the dependent variable and dummy variables for treatments, with L-14 as the omitted category. We also consider a specification that controls for gender, and one that additionally controls for the measured personality characteristics from the questionnaire. The results appear in Table 4. They strengthen the findings from the non-parametric tests: In each specification we find that row 14 choices are significantly less likely to be risky in L-RPS than in L-14, with *p*-Values ranging from 0.041 to 0.026. We also obtain marginal significance of O-14 when controls are added, suggesting a possible framing effect of viewing all 20 decisions. None of the controls themselves are significant predictors of choice. On average women are slightly more risk averse, though the difference is not significant.

If we restrict attention to those subjects that switched only once from Option A to Option B, the results do not change. The percentage choosing the risky option in treatments L-RPS, L-14, and O-14 are 51%%, 72%%, and 56%%, respectively. A chi-squared test and logistic regressions both confirm that the monotonicity failure is statistically significant and the framing effect is marginally significant, with or without controls.

[25] See Siegel and Castellan (1988, Section 8.1) for details on partitioning a chi-squared test in this way.

[26] Additionally, a two-sided Fisher's exact test comparing L-RPS vs. L-14 gives a *p*-Value of 0.061. We chose the chi-squared test *ex ante* because (1) we have three treatments and wanted to use partitioning to provide an independent test of framing, (2) our sample size is adequately large, and (3) the Fisher test is known to be overly conservative (see e.g., Lydersen et al., 2009). Had we planned on using the Fisher test, we would have recruited more subjects to achieve the same power.

Table 5
The three treatments in Experiment 2. The results of S-RPS and S-14 should be equivalent under the assumption of monotonicity. The results of O-14 and S-14 should be equivalent if there is no separated framing effect. The O-14 treatment is identical to Experiment 1 (and previous results are used).

Treatment	S-RPS	S-14	O-14
Decisions (rows) shown	All 20	All 20	Only row 14
Decisions (rows) paid	One random	Only row 14	Only row 14
Display format	Single row	Single row	Single row

Table 6
Percentage of subjects choosing the risky option (Option B) in row 14 in the two separated-format treatments, also compared to the O-14 data from Experiment 1.

Treatment	S-RPS	S-14	O-14 (Experiment 1)
%% Choosing risky	59.0%%	55.6%%	55.7%%
Sample size	61	63	61

We interpret these results as moderate evidence of a monotonicity violation, and weak evidence of a list framing effect. Surprisingly, the two effects almost perfectly cancel out. If we had only compared L-RPS and O-14 (thus confounding incentive compatibility and framing), our regression's treatment effect would have a p-Value of 0.485.

4. Experiment 2: separated decisions

4.1. Design

To test the effect of the list presentation, we created two treatments which are identical to L-RPS and L-14, except each row is shown on a separate screen and the ordering of rows is randomly and independently drawn for each subject. We name these treatments S-RPS and S-14. The S in the treatment names is mnemonic for "Separated display," as opposed to "List display." In both S-RPS and S-14 subjects had to make a choice on each screen before moving on to the next.[27] We recruited 124 new subjects from the same pool as Experiment 1.[28] All other aspects of these treatments were identical to L-RPS and L-14, respectively. We did not re-run treatment O-14 since there is no "list" framing with only one choice. In the analysis that follows, we compare the new S-RPS and S-14 data to our O-14 data from the first experiment.

4.2. Identifying assumptions

The identifying assumptions are detailed in the caption of Table 5. As before, consistency is needed to test the framing effect, while consistency and mechanism invariance are needed to test for monotonicity violations. Here we call the framing effect a *separated framing effect*, to distinguish it from the list framing effect defined above. Comparing O-14 against S-RPS directly (as was done by Harrison and Swarthout, 2014) would require that we assume no such framing effect exists.

Definition 6 (Separated framing effect)**.** A subject exhibits a *separated framing effect* if $\succeq$ differs between an experiment in which only D_i is given and an experiment in which $(D_1, \ldots, D_{20})$ are all given but displayed on separate screens and in random order.

4.3. Results

The percentage of risky choices in S-RPS and S-14 are shown in Table 6, alongside O-14 from Experiment 1. The partitioned chi-squared test does not reject the hypothesis of equal distributions across all three treatments (p-Value 0.9095), and does not reject the hypothesis of monotonicity (p-Value 0.6974).[29] The remaining comparison (O-14 vs. {S-RPS $\cup$ S-14}) is also insignificant (p-Value 0.8444).

Logistic regression results (shown in Table 7) also confirm the treatment differences are not significant, with or without controls.[30] Again, women are slightly more risk averse but the difference is not significant.

We no longer find any differences in choices across these treatments. Thus, when decisions are separated, we find no framing effect and the violations of incentive compatibility disappear.

[27] We also referred to the decision problems as "screens" instead of "rows," to avoid any suggestion that the problems were all part of one list.

[28] The session dates (and number of subjects) were 8/22/2014 (60), 8/25/2014 (52), and 8/26/2014 (12).

[29] Fisher's exact test for monotonicity gives a p-Value of 0.857.

[30] The controls for neuroticism and openness become marginally significant at the 10%% level—both with negative coefficients—but we should expect one or two false positives at this level of significance.

Table 7
Logistic regressions with treatment S-14 omitted. p-Values in parentheses.

Variable	Reg.1	Reg.2	Reg.3
Constant	0.223	0.409	4.224*
	(0.379)	(0.147)	(0.056)
Separated framing effect	0.007	−0.003	−0.119
(Trt. O-14)	(0.984)	(0.993)	(0.752)
Monotonicity violation	0.141	0.157	0.087
(Trt. S-RPS)	(0.697)	(0.366)	(0.381)
Female	×	−0.478	−0.258
		(0.300)	(0.329)
Personality controls	×	×	✓
Either monotonicity violation or Separated framing effect	0.19	0.25	0.29
(F-test of no treatment difference)	0.9095	0.8845	0.8630
Observations	185	185	184
Log likelihood	−126.44	−125.17	−120.11

*10%% significance, **5%% significance, ***1%% significance.

Table 8
Number of switches from Option B back to Option A in the two RPS treatments.

# of B-to-A switches	L-RPS	S-RPS
Zero	95.0%%	67.2%%
One	0%%	29.5%%
Two	0%%	0%%
Three	1.7%%	3.3%%
Four or more	3.3%%	0%%
χ^2 p-Value	*0.00013*	

5. L-RPS and S-RPS as risk measurement devices

Our primary focus is the incentive compatibility of the RPS mechanism. We are not particularly interested in risk elicitation *per se*; the task only serves as a convenient framework in which we can execute our test. But, since we have the data, a natural secondary question is whether the interpretation of subjects' risk preferences would be affected by choice of display format. To this end, we compare S-RPS against L-RPS. But recall that L-RPS was found not to be incentive compatible, so we take the view that the separated format correctly elicits the subjects' underlying risk preferences (under that particular framing), while the list format does not.

We first ask whether switch-back behavior—where a subject chooses Option B on some row k and then Option A on row $k+1$—differs between the two presentation formats.[31] Specifically, we examine whether switch-back behavior becomes more frequent when using the separated format.

The histogram data for the number of switch-backs observed is shown in Table 8.[32] There is a significantly greater frequency of switch-backs in S-RPS, compared to L-RPS (a χ^2 test gives a p-Value of 0.00013). Interestingly, almost all subjects who exhibited switch-backs in S-RPS did so only once. It is tempting to conclude that S-RPS is flawed because it generates more switch-back behavior, but that ignores the fact that S-RPS is incentive compatible while L-RPS is not. A better interpretation is that the L-RPS format generates more consistent *choices*, but we cannot guarantee that underlying *preferences* are in fact more consistent.[33]

We now explore the degree to which switch-back behavior generates inconclusive statements about subjects' risk preferences. Typically, the researcher is interested in identifying the row on which the subject is roughly indifferent between

[31] If we define a preference relation $\sqsupseteq$ over dollar amounts, then we could define monotonicity between $\succeq$ and $\sqsupseteq$ just as we defined it between $\trianglerighteq$ and $\succeq$. If $\sqsupseteq$ is strictly increasing in the dollar amount, then each l_{i+1} dominates l_i ($l_{i+1}(b) \sqsupseteq l_i(b)$ for all b), so monotonicity between $\succeq$ and $\sqsupseteq$ would imply that $l_{i+1} \succ l_i$. But a switch-back between D_i and D_{i+1} leads to the inference that $l_i \succeq l_0 \succeq l_{i+1}$. Assuming transitivity, monotonicity between $\succeq$ and $\sqsupseteq$ is therefore violated. Although this has no implications for incentive compatibility (since it has no implications for the link between $\succeq$ and $\trianglerighteq$), it is still a troubling phenomenon.

[32] Our rate of switch-back behavior in L-RPS (5%%) is a bit low compared to past experiments, but not an outlier. Other examples of low switch-back rates are 2.5%% (Anderson and Mellor, 2009), 5%% (Brown and Kim, 2014), 5.5%% (Holt and Laury, 2002, high stakes), and 10%% (Holt and Laury, 2002, high stakes). On the other extreme, Jacobson and Petrie (2009) found that over 50%% of Rwandan adults exhibited switch-backs, and that those who switch back tend to make worse financial decisions. Similarly high rates of switching back were also observed by Charness et al. (2016). See Charness et al. (2013) for a discussion and survey.

[33] The data from the list display may still provide some useful information about subject preferences. For example, the location of a subject's switch point may correlate with some measure of risk preferences over compound lotteries. But the usual interpretation of choices being truthful would have to be abandoned.

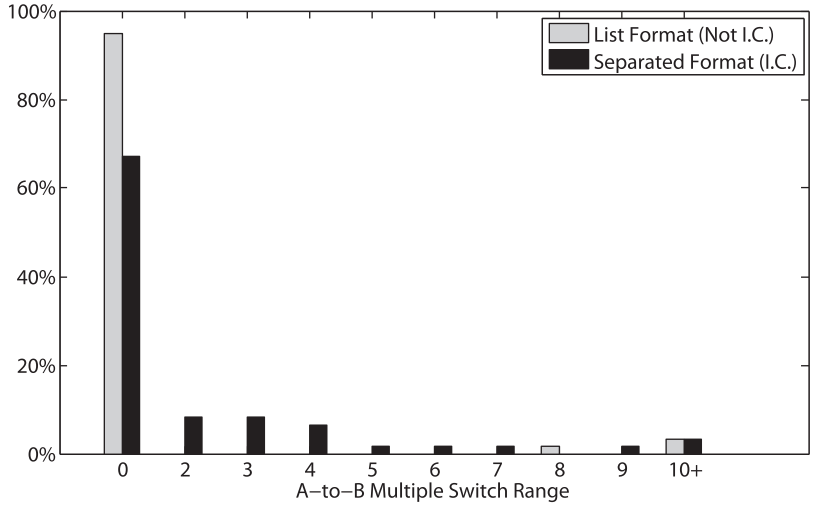

Fig. 3. Histogram of the multiple switch ranges (number of rows between the first and last switch from A to B) by treatment. Zero indicates subjects who only switched once.

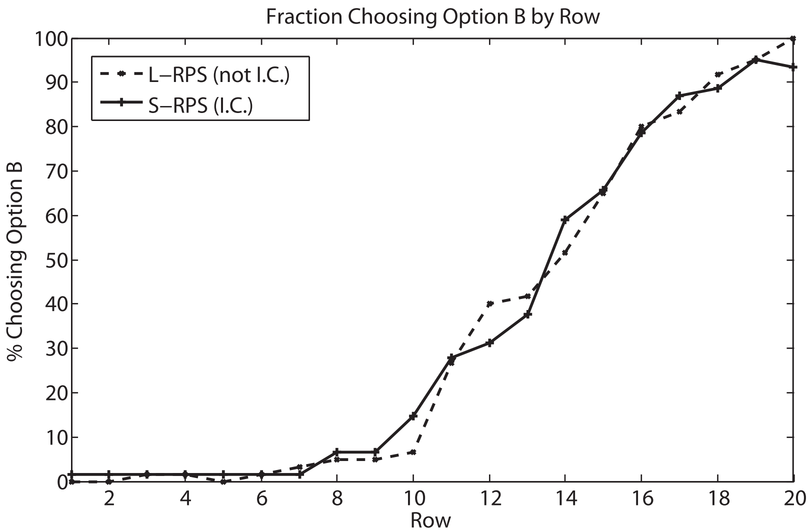

Fig. 4. Frequency of risky choices by row number for the L-RPS (non-incentive compatible) and S-RPS (incentive compatible) treatments.

Option A and Option B. But when a subject switches to B, switches back to A, and then switches again to B, it is often assumed that the "true" indifference row lies somewhere between the first switch to B and the last switch to B. Thus, the larger this interval, the less conclusive the data.

Fig. 3 provides a histogram of the size of these switch-point intervals—the number of rows between the first switch to B and the last—for all subjects. Note that by definition, a subject with consistent preferences has an interval size of zero, and the minimum measure for an inconsistent subject is 2 (a value of 1 is not possible because a subject must switch to A at least once in between the first and last switch to B). Of those subjects who have non-trivial ranges, the majority cover at most 4 rows. This behavior could be rationalized by a model of noisy choice in which the noise only substantially affects choices near the "true" indifference row. This possibility has been discussed by Andersen et al. (2006) and, in the context of value elicitation, by Collins and James (2015). Blavatskyy (2007) provides one such model of noisy choice. A similar interpretation is that subjects have a strict preference for randomization and consciously randomize in rows at which they are nearly indifferent.

Our test of incentive compatibility focuses on choices in row 14. Choice frequencies for all 20 rows are shown in Fig. 4. While not identical, behavior in the two treatments does not appear to differ systematically. Indeed, a χ^2 test confirms that there is no significant difference (p-Value 0.9956). This result should not be particularly surprising: recall that row 14 of L-RPS is similar to O-14 because the list framing effect and the monotonicity violation roughly cancel out. And O-14 is

Table 9
Existing tests of incentive compatibility of the RPS mechanism that have no framing confounds. We describe each of these comparisons in the text below.

Paper	Names of treatments	Presentation format	p-Value	RPS is I.C.?
Starmer and Sugden (1991)	A vs. B	List	0.356	✓
Starmer and Sugden (1991)	C vs. D	List	**0.043**	×
This paper	L-RPS vs. L-14	List	**0.041**	×
This paper	S-RPS vs. S-14	Separated	0.697	✓
Cubitt et al. (1998)	3.1 vs. 3.3	Separated	0.685	✓
Cubitt et al. (1998)	3.2 vs. 3.3	Separated	0.120	✓
Cox et al. (2014b)	PORpi vs. ImpureOT2	Separated[a]	0.122	✓
Cox et al. (2014b)	PORpi vs. ImpureOT3	Separated[a]	0.988	✓
Cox et al. (2014b)	PORpi vs. ImpureOT4	Separated[a]	0.397	✓

[a] Cox et al. (2014b) give subjects the choices on separate slips of paper, but the subjects could have arranged them into a list-like format if they wanted.

similar to row 14 of S-RPS because there is no framing effect and S-RPS is incentive compatible. Thus, L-RPS and S-RPS are similar on row 14. Fig. 4 suggests that this pattern applies to all rows, not just 14.

Of course, this does not mean L-RPS could serve as a convenient proxy for S-RPS. First, we cannot guarantee that the L-RPS and S-RPS will always generate roughly equal choice proportions; their equality here appears more coincidental than structural. The effect of shifting underlying preferences by moving from a separated format to a list format happened to be offset by the distortion caused by the failure of monotonicity. If different lotteries were used, these two effects may no longer offset, leading to significant differences between L-RPS and S-RPS. Second, the apparent similarity in choices at the aggregate level hides differences in behavior at the individual level. In particular, switch-back behavior is significantly greater in S-RPS than L-RPS (see Table 8).

Fig. 4 does reveal a troubling violation of dominance for the S-RPS: subjects who choose Option A in row 20. In that row the risky lottery (Option B) pays \$15 regardless of which ball is drawn. The safe lottery (Option A) is a 50–50 gamble between \$5 and \$10. Any subject who prefers more money to less and for whom $\succeq$ respects monotonicity (with respect to dollar amounts) must choose Option B in row 20.[34] In the L-RPS treatment, none of the 60 subjects choose Option A in row 20. But in the S-RPS treatment, four of the 61 subjects select the dominated option. This value is not significant in our sample with a 2-sided Fisher Exact Test ($p = 0.119$), though we still view these four data points as a possible indicator of confusion or decision fatigue among some of the subjects.

Though we have no direct measure of confusion, we can study the related question of whether subjects experience increasing fatigue as time passes. In particular, we can ask if switch-back behavior in S-RPS is more likely to occur on rows that happened to appear later in the computer interface. In fact, we find the opposite: switch-back behavior is slightly more likely in earlier choices. Thus, if decision fatigue is a factor, it does not appear to increase in time. If anything, it may decrease as subjects become familiar with these lottery choices and computer interface.

6. Related literature

As stated in the introduction, there are several papers that compare an RPS treatment to a One Choice treatment, thus jointly testing both incentive compatibility and framing. Examples include Beattie and Loomes (1997), Cubitt et al. (1998, Experiment 2), Cox et al. (2014a), Freeman et al. (2012), and Harrison and Swarthout (2014). To our knowledge, there are three previously-published papers that compare the RPS treatment to a Framed Control treatment, thus avoiding the confound. They are Starmer and Sugden (1991), Cubitt et al. (1998, Experiment 3), and Cox et al. (2014b). We compare these results alongside ours in Table 9.

Starmer and Sugden (1991) study choices over pairs of lotteries. They compare four treatments, each with 40 subjects. In treatment A (the Framed Control) subjects face decision problems $(D_1, \ldots, D_{20}, D_{21}, D_{22})$ and are paid only for D_{22}. In treatment B (RPS) subjects also face $(D_1, \ldots, D_{20}, D_{21}, D_{22})$ but are paid for D_{21} if a die roll comes up 1, 2, or 3, and paid for D_{22} if the die comes up 4, 5, or 6. Treatments C and D switch the order of the last two decision problems, meaning subjects face $(D_1, \ldots, D_{20}, D_{22}, D_{21})$. In treatment C they are paid for either D_{22} or D_{21}, depending on a die roll, while in treatment D they are paid only for D_{21}. In all four treatments D_{21} and D_{22} were shown on the same page of the subjects' booklets, one right above the other, so we consider these to be presented in a "list" format.

Assuming mechanism invariance, there should be no framing effect when comparing choices in D_{22} between treatments A and B, and also when comparing choices in D_{21} between and C and D. The former comparison yields a p-Value of 0.356,

[34] Again, this is monotonicity between the preference for money and the preference for lotteries and therefore has no consequences for incentive compatibility; see footnote 31.

while the latter's p-Value is 0.043.[35] The conclusion appears to be mixed, but does open some doubt about monotonicity in this setting that uses a list presentation. This is consistent with our results when comparing L-RPS to L-14.

In treatment 3.1 of Cubitt et al. (1998) subjects face binary lottery-choice problems $(D_1, D_2, D_3, \ldots, D_{20})$, each presented on a separate screen and with the order of screens randomized, so we consider this a "separated" presentation format.[36] Treatment 3.1 is a Framed Control, as subjects are paid only for D_1. Treatment 3.2 is another framed control in which subjects are paid only for D_2. Treatment 3.3 is an RPS treatment in which subjects are paid for each decision with a 1/20 probability. Tests of incentive compatibility of the RPS mechanism are obtained by comparing D_1 choice behavior between 3.1 and 3.3, and D_2 choice behavior between 3.2 and 3.3. Neither gives significant differences, with p-Values of 0.685 and 0.120, respectively. Thus, with the separated presentation we find no evidence of incentive compatibility failures for the RPS mechanism. This is consistent with our results when comparing S-RPS to S-14.

Cox et al. (2014b) compare behavior across an impressive number of different payment mechanisms, and find significant differences in behavior between them. Their subjects face five binary lottery-choice problems $(D_1, \ldots, D_5)$, each presented on a separate slip of paper. The five slips of paper are placed in random order into an envelope which is then given to the subject. We label this presentation format as separated, but note that subjects could order the slips of papers as a list if they so choose. Their treatments include "PORpi", which is an RPS treatment; ImpureOT2, which is a Framed Control that pays only for the second slip of paper; ImpureOT3, which is a Framed Control that pays only for the third slip of paper; and ImpureOT4, which is a Framed Control that pays only for the fourth slip of paper. Comparing decisions in PORpi to each of the ImpureOT treatments gives a clean test of incentive compatibility of the RPS mechanism, assuming mechanism invariance. Using χ^2 tests, we find no significant differences between the RPS treatment and the various ImpureOT treatments, suggesting incentive compatibility of the RPS mechanism was satisfied in their experiment.[37] This conclusion is maintained even when the three ImpureOT treatments are pooled (p-Value 0.643).

Sadiraj and Sun (2012) test the incentive compatibility of the RPS mechanism in a game-theoretic setting. Subjects engage in six bargaining games against different partners. Treatments vary on whether subjects bargain over gains and losses, and whether the RPS mechanism is used or only the fifth round is paid with certainty. Probit regressions confined to the fifth round show a significant effect of both treatment variations, suggesting that bargaining differs between gains and losses, and that the RPS mechanism may not be incentive compatible in this setting. We interpret this finding with some caution, however, because behavior in rounds one through four were different across payment mechanisms and this could lead to difference preferences in round five.

Biases in list elicitation procedures are explored by Andersen et al. (2006), Beauchamp et al. (2012), Sprenger (2015), Castillo and Eil (2013), Kim and Rosenblat (2015), Zuo and Zhang (2015), and several others. The idea of scrambling or randomizing the rows of a multiple price list is not novel. Kirby and Marakovic (1996) estimate discount factors of college students using a scrambled list of binary choices shown on one page, and Kirby et al. (1999) repeat this procedure with heroin addicts. Eckel et al. (2005) elicit time preferences of the working poor using a scrambled multiple price list with decisions shown separately, as we do here. In a footnote, they say "we now believe that scrambling is a bad idea because it results in greater inconsistency and variance of responses." Our results suggest the opposite conclusion: presenting decisions in a list format may be a bad idea because it might hide subjects' true inconsistency and variance of responses.

7. Discussion

We provide evidence which calls into question the assumption that choices made by subjects in a list setting actually reflect underlying preferences. This violation of incentive compatibility appears to be restored when these decisions are separated and presented to subjects in random order. The previous literature supports this conclusion, as Table 9 reveals.

Based on these conclusions we suggest that researchers should consider using the separated format in elicitation tasks, as well as other types of experiments with multiple decisions. This does force researchers to deal with less consistent choice data, such as switch-back behavior in list elicitation procedures. A promising avenue of future research would be to study the source of these inconsistencies. Our conjecture is that switch-backs occur near indifference, and that subjects may tremble frequently on such decisions. Most switch-back regions are not particularly large, supporting this interpretation of results. Further studies on the stochasticity of choice—including a careful definition of incentive compatibility under stochastic choice—would be insightful.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.euroecorev.2017.09.014.

[35] Starmer and Sugden pool treatments B and C together in their analysis because there are no significant differences between them. The comparison of A to B∪C yields a difference that is just barely insignificant at standard levels, with a p-Value of 0.051, while differences between D and B∪C are not significant, with a p-Value of 0.14.

[36] Subjects can backtrack at any point and change any prior decisions before submitting.

[37] These tests do not appear in their paper; we thank Cox, Sadiraj and Schmidt for sharing their data.

References

Andersen, S., Harrison, G.W., Lau, M.I., Rutstrom, E.E., 2006. Elicitation using multiple price list formats. Exp. Econ. 9, 383–405.
Anderson, L.R., Mellor, J.M., 2009. Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. J. Risk Uncertain. 39, 137–160.
Azrieli, Y., Chambers, C.P., Healy, P.J., 2012. Incentives in Experiments with Objective Lotteries. Ohio State University. Working Paper.
Azrieli, Y., Chambers, C.P., Healy, P.J., 2016. Incentives in Experiments: A Theoretical Analysis. Ohio State University. Working Paper.
Bade, S., 2012. Independent Randomization Devices and the Elicitation of Ambiguity Averse Preferences. Max Plank Institute. Working Paper.
Baillon, A., Halevy, Y., Li, C., 2014. Experimental Elictation of Ambiguity Attitude Using the Random Incentive System. University of British Columbia. Working Paper.
Beattie, J., Loomes, G., 1997. The impact of incentives upon risky choice experiments. J. Risk Uncertain. 14, 155–168.
Beauchamp, J.P., Benjamin, D.J., Chabris, C.F., Laibson, D.I., 2012. How Malleable are Risk Preferences and Loss Aversion. Harvard University. Working Paper.
Becker, G.M., DeGroot, M.H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behav. Sci. 9, 226–232.
Blavatskyy, P.R., 2007. Stochastic expected utility theory. J. Risk Uncertain. 34, 259–286.
Brown, A.L., Kim, H., 2014. Do individuals have preferences used in macro-finance models? An experimental investigation. Manag. Sci. 60, 939–958.
Castillo, M., Eil, D., 2013. A Two-Way Street: Multiple Price Lists, the Common Ratio Effect, and Preference Reversals. George Mason University. Working Paper.
Charness, G., Gneezy, U., Imas, A., 2013. Experimental methods: eliciting risk preferences. J. Econ. Behav. Organ. 87, 43–51.
Charness, G., Viceisza, A., et al., 2016. Three risk-elicitation methods in the field: evidence from rural senegal. Rev. Behav. Econ. 3 (2), 145–171.
Collins, S.M., James, D., 2015. Response mode and stochastic choice together explain preference reversals. Quant. Econ. 6, 825–856.
Cox, J., Sadiraj, V., Schmidt, U., 2014a. Asymmetrically dominated choice problems, the isolation hypothesis and random incentive mechanisms. PLoS One 9 (3), e90742.
Cox, J., Sadiraj, V., Schmidt, U., 2014b. Paradoxes and mechanisms for choice under risk. Exp. Econ. 18 (2), 215–250.
Cubitt, R.P., Starmer, C., Sugden, R., 1998. On the validity of the random lottery incentive system. Exp. Econ. 1, 115–131.
Eckel, C., Johnson, C., Montmarquette, C., 2005. Savings decisions of the working poor: short- and long-term horizons. In: Harrison, G.W., Carpenter, J., List, J.A. (Eds.), Field Experiments in Economics. In: Research in Experimental Economics, vol. 10. Emerald Group Publishing, pp. 219–260.
Freeman, D., Halevy, Y., Kneeland, T., 2012. Probability List Elicitation for Lotteries. University of British Columbia. Working Paper.
Grether, D.M., Plott, C.R., 1979. Economic theory of choice and the preference reversal phenomenon. Am. Econ. Rev. 69, 623–638.
Harrison, G.W., Swarthout, J.T., 2014. Experimental payment protocols and the bipolar behaviorist. Theory Decis. 77, 423–438.
Holt, C.A., 1986. Preference reversals and the independence axiom. Am. Econ. Rev. 76, 508–515.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. Am. Econ. Rev. 92 (5), 1644–1655.
Jacobson, S., Petrie, R., 2009. Learning from mistakes: What do inconsistent choices over risk tell us? J. Risk Uncertain. 38, 143–158.
Kahneman, D., 1992. Reference points, anchors, norms, and mixed feelings. Organ. Behav. Hum. Decis. Process. 51, 296–312.
Kahneman, D., Tversky, A., 1984. Choices, values, and frames. Am. Psychol. 39, 341–350.
Karni, E., Safra, Z., 1987. "Preference reversal" and the observability of preferences by experimental methods. Econometrica 55 (3), 675–685.
Kim, Y., Rosenblat, T.S., 2015. Reference-Dependent Preferences in Risk Preference Elicitation Methods With a Multiple Price List Format. University of Nebraska. Working Paper.
Kirby, K.N., Marakovic, N.N., 1996. Delay-discounting probabilistic rewards: rates decrease as amounts increase. Psychon. Bull. Rev. 3, 100–104.
Kirby, K.N., Petry, N.M., Bickel, W.K., 1999. Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. J. Exp. Psychol.: Gen. 128, 78–87.
Kuzmics, C., 2017. Abraham Wald's complete class theorem and Knightian uncertainty. Games Econ. Behav. 104, 666–673.
Lydersen, S., Fagerland, M.W., Laake, P., 2009. Recommended tests for association in 2x2 tables. Stat. Med. 28, 1159–1175.
Machina, M.J., Schmeidler, D., 1992. A more robust definition of subjective probability. Econometrica 60 (4), 745–780.
Oechssler, J., Rau, H., Roomets, A., 2016. Hedging and Ambiguity. University of Heidelberg. Working Paper.
Oechssler, J., Roomets, A., 2014. Unintended hedging in ambiguity experiments. Econ. Lett. 122 (2), 243–246.
Sadiraj, V., Sun, J., 2012. Efficiency in bargaining games with alternating offers. Econ. Bull. 32 (3), 2366–2374.
Safra, Z., Segal, U., Spivak, A., 1990. Preference reversal and nonexpected utility behavior. Am. Econ. Rev. 80, 922–930.
Savage, L.J., 1954. The Foundations of Statistics. John Wiley & Sons, New York, NY.
Segal, U., 1988. Does the preference reversal phenomenon necessarily contradict the independence axiom? Am. Econ. Rev. 78, 233–236.
Segal, U., 1990. Two-stage lotteries without the reduction axiom. Econometrica 58 (2), 349–377.
Siegel, S., Castellan Jr., N.J., 1988. Nonparametric Statistics for the Behavioral Sciences, second ed. McGraw-Hill, New York, NY.
Sprenger, C., 2015. An endowment effect for risk: Experimental tests of stochastic reference points. J. Polit. Econ. 123 (6), 1456–1499.
Starmer, C., Sugden, R., 1991. Does the random-lottery incentive system elicit true preferences? An experimental investigation. Am. Econ. Rev. 81, 971–978.
Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211 (4481), 453–458.
Von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior, third ed. Princeton University Press, Princeton, NJ.
Yaari, M.E., 1965. Convexity in the theory of choice under risk. Q. J. Econ. 79 (2), 278–290.
Zuo, S.X., Zhang, Y.J., 2015. Multiple Switches in the Multiple Price List: Confused or Inconsistent?. University of Houston. Working Paper.