

Salient Rewards, Payoff Protocols and Biased Data

James C. Cox and Vjollca Sadiraj

Experimental Economics Center and Department of Economics

Georgia State University

July 2026

Keywords: Salient payoffs, Payoff protocols, Incentive compatibility, Random selection, Experimental procedures

Forthcoming in *Southern Economic Journal* Special Issue in Honor of James C. Cox

Abstract

Paying salient rewards to subjects is expensive. Is it worth it? We describe previous data from otherwise identical treatments with hypothetical or salient payoffs that indicate opposite conclusions about the central research question in a paper. If we are going to pay salient rewards, how should we? The answer depends on the theoretical model underlying hypotheses being tested with the data, hence the choice of payoff protocol is an essential feature of experimental design. In recent years there has been renewed interest in theoretical and experimental research on properties of payoff protocols. We present an extension and critical discussion of the literature intended to promote uptake.

1. Introduction

The dominant view among experimental economists is that researchers should pay salient rewards to subjects in experiments. There is revealed disagreement about the specifics of saliency: how to relate subject rewards to experiment outcomes to incentivize unbiased choice from options in the experiment. We discuss “should”, “salient” and “how”.

Why should researchers pay salient rewards? One answer is that data from experiments without salient rewards can be inconsistent with *economically motivated* behavior, the central topic of economics. There is recent literature on the statistical significance or insignificance of differences in average (or some other statistic) outcomes with hypothetical vs. salient rewards (Brañas-Garza et al. 2023; Hackethal et al. 2023). We bring back to the attention of the profession that subjects’ behavior elicited with hypothetical rewards has been observed to support the *opposite* conclusion about the preference reversal phenomenon than does behavior elicited with salient rewards. And we discuss the characteristics of subjects’ tasks that produced this outcome.

In a seminal paper, Smith wrote: “... *In both the field and the laboratory, value is induced on messages by the institution* whose rules state how messages are to be translated into valuable outcomes. ... Thus in an experiment, ... we must also give the subject a property right to rewards that are related appropriately to the realized experimental outcomes” (Smith, 1982, pg. 931). In order for a subject’s messages to be related “appropriately” to rewards, the institution must be incentive compatible: it must always give a subject whose most preferred outcome is associated with message M an incentive to choose M rather than some other feasible message. The relationship between this last straightforward statement and much experimental practice is more

interesting and more controversial than it might at first appear to be. We explore this topic in our paper and argue that the answers about incentive compatibility vary with the purpose of the experiment and the theory being applied or tested. There is no unique payoff protocol that is appropriate for all research questions. The choice of payoff protocol should be treated as an essential feature of experimental design. It often is not, as made clear by two surveys of the literature reported in section 7. Section 4 focuses on theoretical results, new and old, on incentive compatibility of pay-one-randomly payoff protocol¹, an often-used payoff protocol and the very one that inspired thought-provoking ideas in the preference reversal literature forty years ago. Section 5 focuses on recent discussions of the empirical properties of pay-one-randomly. Section 6 reports on incentive compatibility of alternative payoff protocols for salient payoffs. Section 8 offers suggestions for experimental methods.

We begin section 2 by discussing some content in the literature on the preference reversal phenomenon which, forty years ago, first focused researchers' attention on potential effects of payoff protocols on individual choices and valuations, that when not accounted for, may lead to incorrect inferences pertaining to the original research question of a study.

2. Important lessons from preference reversal experiments

Preference reversal experiments originated in psychology and were introduced into the economics literature by Grether and Plott (1979).² Conventional preference reversal experiments use pairs of binary lotteries for which subjects are asked to: (1) choose between the lotteries in each pair; and (2) state minimum selling price for each lottery in each pair using BDM mechanism (Becker, et al. 1964). Each pair includes a “dollar bet” with a relatively low probability of a relatively high win-state payoff, and a “probability bet” with a relatively high probability of a relatively low win-state payoff. Many subjects state higher selling price for the dollar bet than for the probability bet but choose the latter over the former; this is known as “predicted” preference reversal. An “unpredicted” reversal is when a subject chooses the dollar bet but places a higher selling price on the probability bet. A common feature of data from preference reversal experiments is much higher

¹ This payoff protocol has been given several names in the literature including random lottery incentive mechanism (RLIM), random incentive system (RIS), random incentive mechanism (RIM), random problem selection (RPS) and pay one randomly (POR).

² The standard interpretation of the preference reversal phenomenon by psychologists is inconsistent with the conventional economics view that preferences are a characteristic of an agent. Instead, a common view in psychology is that the preferences of an agent are dependent on the response mode used to elicit them (Slovic and Lichtenstein 1968; Lichtenstein and Slovic 1971).

incidence of predicted reversals than unpredicted reversals.³ Psychologists argue this can be explained by the anchoring and adjustment theory (and later variations).⁴ But if risk preferences are a characteristic of a rational agent then elicited preferences should be independent of response modes and the higher stated price should be for the chosen lottery.

2.1 Hypothetical and salient incentives with BDM and random decision selection

Grether and Plott (1979) were skeptical about the preference reversal experiments reported by psychologists for reasons including their use of hypothetical payoffs. A central feature of the Grether and Plott design was use of paired treatments, one with hypothetical payoffs and another (otherwise identical) treatment with salient monetary payoffs. They report *higher* incidence of predicted reversals in the monetary payoff treatment than in the hypothetical payoff treatment. Grether and Plott (1979, pg. 623) concluded that “...this behavior is not simply a violation of some type of expected utility hypothesis. The preference measured one way is the *reverse* of preference measured another and seemingly theoretically compatible way. ... The inconsistency is deeper than mere lack of transitivity or stochastic transitivity.”

Grether and Plott’s central conclusion was challenged by Holt (1986) and Karni and Safra (1987) who maintained that the observed behavior could indeed be simply a violation of the expected utility hypothesis. Holt (1986) shows that, given reduction of compound lotteries, Grether and Plott’s procedure of randomly selecting one of a subject’s multiple responses for payoff depends on the independence axiom of expected utility theory to interpret the data as revealing preference reversals. Karni and Safra (1987) explained that use of the BDM procedure to elicit responses that could be interpreted as (minimum) selling prices required the reduction and independence axioms. These criticisms introduced an important principle into the methodology of experimental economics.

Remark 1. The validity of conclusions that can be drawn from experiment data depends on the elicitation method and the payoff protocol. The choice of payoff protocol is an essential feature of experimental design.

³ Note that simply including unbiased random errors in decision theory cannot explain higher relative frequencies of predicted than unpredicted reversals.

⁴ See Lichtenstein and Slovic (1971) and Tversky et al. (1990). For an informative survey see Seidl (2002). Sugden (2015) presents a discussion of the psychology of “inner rational agent”.

Forty years later, this lesson has still not been internalized in the practice of experimental economics: see section 7 below.

2.2 Hypothetical and salient incentives without BDM or random decision selection

Within the preference reversal literature, Cox and Epstein (1989) and Cox and Grether (1996) responded to the methodological critique by conducting experiments that do *not* use either random decision selection or BDM. These researchers realized payoffs after each round and paid each subject the total of their payoffs at the end of an experiment session.⁵ The data was analyzed for possible income effects, and for order effects, which were not found; hence their data is usable for studying response mode effects.

2.2.1 Opposite conclusions from hypothetical and salient incentives in repetitive markets

In the first round in the market treatments, significant asymmetric (predicted greater than unpredicted) preference reversals were observed with *both* hypothetical and salient payoffs. By round 5, the outcomes changed dramatically. With hypothetical payoffs, significant asymmetric preference reversals were robust to five rounds of repetition in the markets. In contrast, with salient payoffs, predicted reversals were slightly lower than unpredicted reversals after five market rounds. In this way, data from hypothetical or salient payoff treatments lead to opposite conclusions about the central question posed in the research: Is the *predicted* reversals pattern robust to repetitive decisions in markets? This provides an example of unambiguously clear support for the importance of using salient incentives in experiments to study *economic* behavior.

Remark 2. Data from hypothetical vs. salient payoff treatments led to opposite conclusions regarding prevalence of predicted preference reversals.

2.2.2 What accounts for the contradiction between results from hypothetical and salient payoffs?

The general idea is that with hypothetical payoffs there are no monetary costs from making suboptimal choices, and therefore there are weak incentives for learning. Specific to this study, there are some distinctive features of the tasks. The market (sequential) choice task is the English clock auction in which, at each tick of the price clock, a subject chooses between keeping the lottery for sure (by exiting the auction) or maintaining the right to sell it. Lottery pair 2 (in Cox

⁵ See section 6 below for incentive compatibility properties of this “pay all sequentially” payoff protocol.

and Grether 1996) provides a representative example.⁶ In the English auction, clock price starts at the amount of the win-state payoff and decreases by 5 cents with each tick, every second. In both hypothetical and incentivized treatments, in the dollar bet the clock price starts at \$9 whereas in the probability bet it starts at \$2. For the dollar bet it takes 153 ticks, extending over 153 seconds, for the price on the clock to decrease from the \$9 starting price to the bet's expected value of \$1.35. Subjects may be impatient or get disutility from the attention-intensive activity of watching the price clock tick down and choosing at each tick whether to exit. If so, they can be relieved of these costs by exiting the auction. Of course, exiting early produces high minimum selling prices for the dollar bet and results in high predicted preference reversal rates. This is what many subjects did with hypothetical payoffs but not with salient payoffs. On average, subjects exited the English auction for the pair 2 dollar-bet 74 seconds earlier (at a \$3.7 higher price) with hypothetical payoffs than with salient payoffs. This central finding in Cox and Grether (1996) clearly shows why it can be important to *avoid* hypothetical payoffs.

The early critiques of experimental methods in the preference reversals literature apply whether there are or are not response mode effects that can elicit different risk preferences. We now turn our attention towards alternative payoff protocols used in many-decision experiments for eliciting preferences that are a characteristic of an agent.

3. Alternative mechanisms for salient payoffs

The Holt (1986) critique of Grether and Plott (1979) makes clear the importance of incorporating properties of payoff protocols in hypotheses and data analysis for drawing valid conclusions from experiments. Such questions about payoff protocols arise in many types of experiments.⁷ Our focus will be on experiments on decision making under risk.

Discussion of experimental methods for researching decisions under risk is dependent on an answer to the seminal question about what the researcher hopes to learn from conducting an experiment.⁸ Is the purpose of an experiment to elicit subjects' latent risk preferences? Or is the purpose of an experiment to study how subjects respond to decision environments that provide

⁶ In lottery pair 2: the dollar-bet offers 7 chances to win \$9 and 29 chances to lose \$0.50; the probability-bet offers 29 chances to win \$2 and 7 chances to lose \$0.50.

⁷ Examples include experiments on decision-making under risk or ambiguity, trust (investment) games with role reversal, public good games, voting and bargaining with alternating offers (Cox et al. 2015).

⁸ Cox and Sadiraj (2019) characterize the relationship between an experimenter and subjects as a principal-agent problem. They discuss several examples of misalignment between the objectives of experimenter and subjects.

opportunities for constructing risk preferences? We will be concerned with experimental methods for eliciting latent risk preferences.

Suppose a researcher wants to elicit a subject's preference over two options. The traditional "gold standard" method is to ask the subject to choose between the two options and pay the realized outcome from the chosen one. This single task elicitation is incentive compatible for latent preferences and it is used in experiments to test hypotheses with between-subjects data.

Remark 3. Single task elicitation is incentive compatible for eliciting latent risk preferences. It requires between-subjects designs.

There are many experiments that use within-subjects data generated by having individual subjects make multiple decisions. In such experiments, payment of salient rewards necessitates choice of a protocol or "mechanism" for rewarding subjects. In recent years there has been renewed interest in theoretical and experimental research on incentive compatibility of payoff protocols. Significant progress has been made, however seemingly-inconsistent statements of results in some papers makes uptake by readers of the current state of knowledge a difficult endeavor. We report theoretical results on incentive compatibility and integrate results from recent literature on payoff protocols to promote uptake. We begin with a critical discussion of theoretical and recent empirical literature on pay-one-randomly because: (i) it was the focus of critical discussion of implications of preference reversals for rationality forty years ago; (ii) it is the protocol that has received special attention in recent literature; and (iii) it is the most used protocol in experiments on individual decision making under risk. This payoff protocol has been given several names;⁹ we use "pay-one-randomly".

4. Theory of pay-one-randomly incentives

4.1 Popular Decision Theories

Randomly selecting one of a subject's decisions for payment is a lottery. When choice options in an experiment are over simple lotteries, use of pay-one-randomly to pay subjects creates compound lotteries. Previous literature (e.g. Holt 1986; Cox et al. 2015) demonstrates that pay-one-randomly is incentive compatible for expected utility theory but is *not* incentive compatible for theories with nonlinear probability weights, such as rank dependent utility theory (Quiggin

⁹ The several names for this payoff protocol are listed in footnote 1 above.

1982), cumulative prospect theory (Tversky and Kahneman 1992), and dual theory of expected utility (Yaari 1987).¹⁰ Given reduction of compound lotteries, but absent independence, randomly selecting one decision for payoff creates incentives for biased revelation of latent risk preferences. Recent work by Azrieli et al. (2020) also explains that, given reduction, pay-one-randomly is not incentive compatible for rank dependent utility, in this way reaching the same conclusion as Cox et al. (2015).

Remark 4. Given reduction, pay-one-randomly is incentive compatible for expected utility theory but not for rank dependent utility theory, cumulative prospect theory, or dual theory of expected utility.

4.2 Stochastic Choice

We draw attention to the key idea that pay-one-randomly protocol provides incentives for stochastic choice for subjects with preferences for mixtures (Machina 1985). Stochastic choice (different choices in repeated decision tasks with the same feasible set) is not a thought experiment, it is an empirically prevalent phenomenon.¹¹

Consider two non-degenerate lotteries, $\{A, B\}$ and a feasible set, $\zeta = \{\alpha, \beta, \gamma\}$ such that option α offers lottery A with certainty, option β offers lottery B with certainty, and option γ offers A or B with equal probability. Our individual strictly prefers lottery A over B but also strictly prefers an equal randomization of A and B against getting either one for sure. As commonly assumed in theoretical literature, our individual is indifferent between a lottery and the degenerate distribution that offers the lottery with certainty. It follows that our individual's ranking of options in ζ is $\alpha > \beta$, and $\gamma > \beta$ and $\gamma > \alpha$.

The experimenter is interested in eliciting this individual's ranking of A and B. In a single-task, pay-all-subjects experiment, the experimenter promises to pay the chosen option with probability 1. So, when our individual is asked to choose between lotteries A and B, she actually chooses between getting A with certainty or getting B with certainty. That is, her actual feasible

¹⁰ E.g. Holt 1986, Cox et al. 2015.

¹¹ Agranov and Ortoleva (2017) report that in their experiment the majority of subjects choose different lotteries when asked to choose from the same set of lotteries three times in a row and are told about the repetition. For an overview of empirical literature on preferences for randomization see Agranov and Ortoleva (2022).

set is $\{\alpha, \beta\}$, and given her preferences, $\alpha \succ \beta$, she chooses α , and the experimenter observes choice A.

Now, consider our individual being asked to choose between lotteries A and B twice in a two-decision-task experiment with pay-one-randomly incentives. In this two-decision task experiment, her feasible set is $\zeta = \{\alpha, \beta, \gamma\}$. The most preferred option in ζ for our individual is γ , and the only way she can get that is by making a stochastic choice, choose A from $\{A, B\}$ in one of the decision tasks, and B from the same set $\{A, B\}$ in the other decision task. That is, as Machina (1985) reasons, even though our individual has deterministic preferences over A and B, when offered the opportunity to choose more than once from $\{A, B\}$ her choice will be stochastic.

Remark 5. Pay-one-randomly is not incentive compatible for agents with preferences for randomization.

5. Empirical properties of pay-one-randomly protocol

In this section we discuss biased elicitation by the pay-one-randomly protocol, the most used protocol in experiments on individual decision making under risk. Use of pay-one-randomly was defended, initially, by appeal to the isolation hypothesis and, recently, by assuming statewise monotonicity with decisions identifying states (Azrieli et al. 2018).

5.1 Empirical validity of the isolation hypothesis

There has been accumulating evidence that: (a) decision makers do *not* generally isolate individual decisions; and (b) bias from using pay-one-randomly can be significant. We integrate and interpret findings reported in several recent papers.

5.1.1 Isolation vs. whole experiment optimization

Earlier studies (Starmer and Sugden 1991; Cubitt et al. 1998) provide data that favor the isolation hypothesis over the alternative hypothesis of “full reduction”, which means subjects make every choice so as to yield the most preferred compound lottery from *the whole* experiment.¹² But full reduction vs. isolation are extreme hypotheses. A more plausible alternative to isolation is provided by cross-task contamination in which choices may be affected by options included in the adjacent decision tasks. Starmer and Sugden (1991, p.977) data suggest cross-task contamination. In

¹² See also Hey and Lee (2005a, 2005b) for related experiments.

contrast, Cubitt et al. (1998) do not find cross-task contamination from random task selection when choice options are presented in list format. This provides a bridge to subsequent literature.

Remark 6. Full reduction is rejected in favor of isolation.

5.1.2 Isolation vs. cross-task contamination

Cox et al. (2015) report data from elicitation of choices from five classic lottery pairs¹³ using nine payoff protocols (or mechanisms) including two alternative versions of pay-one-randomly and two versions of single-task salient payment treatment. The data are dramatically inconsistent with the isolation hypothesis. The isolation hypothesis requires no payoff protocol effects on the percentages of the less risky option being chosen in each of the five lottery pairs. For three (out of the five) lottery pairs, the frequency of the less risky lottery choice in some protocol is more than three times the figure in a different protocol whereas for the remaining two lottery pairs it is more than twice.¹⁴ Other recent papers that report data inconsistent with the isolation hypothesis include Cox et al. 2014, Harrison and Swarthout 2014, Brown and Healy 2018, Charness et al. 2018, Freeman and Mayraz 2019 and Freeman et al. 2019. Some of these papers provide evidence on potential sources of biased choices with pay-one-randomly: incentives, list formats, menu dependence, and certainty effects.

Remark 7. Isolation is rejected in favor of cross-task contamination.

5.2 Observed Behavior with Pay-One-Randomly

Pay-one-randomly can affect elicited risk preferences in at least three distinct ways. One comes from the incentives created: the interpretation of data requires a theory of how subjects respond to compound lotteries. As stated in section 4, a subject whose preferences are characterized by the commonly used non-expected utility theories may be incentivized to reveal biased preferences in an experiment using pay-one-randomly. An immediate implication is that data from the large literature reporting experiments using pay-one-randomly, intended to test hypotheses from cumulative prospect theory or rank dependent utility theory, could be biased for *the decision theory applied to the data*. The researchers who use pay-one-randomly for testing or applying non-EU

¹³ All treatments involved choices over the same five pairs of binary lotteries that had the potential for subjects to make choices exhibiting Allais (1953) paradoxes and dual Yaari (1987) paradoxes.

¹⁴ See Table 4 in Cox et al. (2015).

decision theories need to show that their data are not biased; including single-task controls for subsets of the decision tasks in the experiment, as in Harrison and Swarthout (2014), seems promising.

A second way in which pay-one-randomly can affect elicited choices comes from an experimenter's decision about the set of options included in the experiment as determined by the research questions of the study. Different compositions of this menu of choices can cause pay-one-randomly to elicit different preferences over individual choice pairs. For example, inclusion or exclusion of asymmetrically dominated lotteries in the menu of options has been shown to elicit significantly different risk preferences (Cox, et al. 2014).

A third way in which pay-one-randomly can affect elicited choices occurs if a subject's preferences are characterized by the certainty effect (Allais 1953). With such preferences, use of pay-one-randomly will transform a certain payoff (a degenerate lottery) into a stochastic payoff received with probability $1/n$, where n is the number of decision tasks in the experiment. This can cause elicited preferences over a certain prize and a nondegenerate lottery to be the reverse of preferences elicited by the single-task protocol in which the certain option is received with probability 1 when chosen by the subject (Freeman et al. 2019).¹⁵

Recent studies (Brown and Healy 2018; Freeman and Mayraz 2019) focus on disentangling incentives from presentation effects such as: (a) making one choice from a list; (b) making sequential choices from a list; (c) seeing a list and then making one choice on a separated screen; and (d) not seeing a list but sequentially making choices on separated screens. One of the objectives of these papers is to develop experimental procedures intended to minimize bias in data generated through use of pay-one-randomly to pay subjects. Such developments have the potential to offer researchers valuable tools. We discuss some researched procedures in individual papers (Table 1) and evaluate them in the larger context of other papers in recent literature to ascertain the robustness of reported findings and conclusions.

<Insert Table 1 about here>

5.2.1 Menu effects

A menu effect occurs when the choice among options in one task can depend on the options available in other tasks. Cox, et al. (2014), who experiment with single-task and two tasks with

¹⁵ Note that this feature is always present in one decision task experiments with random *subject* payment protocol.

separated decisions and pay-one-randomly¹⁶, find that choices depend on presence or absence of asymmetrically dominated choice options in the experiment. This is a menu effect.

Freeman and Mayraz (2019) experiment with R-lists, K-lists, and SCs. R-lists are treatments which pay subjects according to pay-one-randomly from options presented in lists (Holt and Laury 2002). Presentations of lottery pairs in K-lists are the same as in R-lists but the subjects are informed they will be paid according to their choice in one *known* pair (hereafter “pay-one-known”) rather than a randomly selected choice pair. SC treatments offer subjects a *single* choice pair (i.e., single-task). They report that their data indicate that single-task elicits different choices than pay-one-known (K-list, row 11, Table 1) which has the same financial incentives, and also different choices than pay-one-randomly (R-list, row 10, Table 1). This naturally calls for examining whether separated decisions can be a remedy for the menu effects.

5.2.2 Separated choice format effects

Brown and Healy (2018) examines behavior when decision tasks are presented separately (S) or as a list (L). One of their findings is that with lists, pay-one-randomly (L-RPS) elicits more risk aversion than pay-one-known task (L-14)¹⁷ leading to the conclusion that pay-one-randomly lists elicits biased data (row 1, Table 1).

They attribute the bias to the list inducing menu-dependent preferences as, absent lists, in their experiment choices in pay-one-randomly (S-RPS) and pay-one-known task (S-14) are similar (row 2, Table 1).¹⁸ They conclude that when decision tasks are separated (for example, each displayed on a separate computer screen), pay-one-randomly does not elicit significant bias. This is consistent with Cox et al. (2015) data in treatments pay-one-randomly with prior information (PORpi) vs. pay-one-known decision (ImpureOT) shown in row 3, Table 1.

These results suggest there might be a simple solution to the problem of pay-one-randomly producing biased data: present choice tasks on separate screens rather than in list format. This would be important for experimental methods if it were a robust result. Unfortunately, this simple

¹⁶ All decision tasks were given to the subjects in the instructions.

¹⁷ Brown and Healy 2018, Table 3: The risky option in decision 14 is chosen by 51.7% (list pay-one-randomly L-RPS), and increases to 70% (list pay-decision-14, L-14) of subjects. Options in their decision task 14 are: 10 or 5 (equally likely) vs 15 with probability 0.7.

¹⁸ Brown and Healy 2018, Table 6. The observed frequencies of the risky option in decision 14 are 59% (separated pay-one-randomly, S-RPS) and 55.6% (separated, pay-decision-14, S-14).

remedy is not supported by data from Harrison and Swarthout (2014), Freeman et al. (2019), and data from other treatments in Cox et al (2015).

Harrison and Swarthout (2014) compares risk attitudes estimated for rank dependent utility model with data from the single-task treatment in which subjects make a single choice (treatment “1 in 1”) and are paid for it and data from a treatment with *separated* decisions in which subjects make 30 choices and are incentivized with pay-one-randomly (treatment “1 in 30”) protocol.¹⁹ Their data reject the null hypothesis that “1 in 1” and “1 in 30” choices can be characterized by the same risk aversion parameters for rank dependent utility model (Table 1, row 12).

In Cox, et al. (2015) treatment with pay-one-randomly and no prior information (PORnp) the first time a subject saw a lottery pair was when they encountered the choice screen for that pair.²⁰ The significantly lower risk aversion observed in single-task than in pay-one-randomly, no prior information treatment (using all rounds or only the first-round choices) cannot be attributed to a list format (Table 1, rows 4 and 5).

Freeman et al. (2019) data in treatments with certainties also do *not* provide support for eliminating bias from pay-one-randomly by presenting choice tasks on separate screens (Table 1, row 8). They attribute the bias to certainty effects, the topic of our next section.

5.2.3 Certainty effects

Pairwise choices in list format and pay-one-randomly are often used to elicit certainty equivalents of lotteries (in place of BDM). Each row in the list contains the same lottery but the certain amount of money changes (usually monotonically) across rows. If, instead, a switching probability is to be elicited, each row in the list contains the same amount of money but probability of the outcomes (of the risky option) monotonically changes across rows.

Freeman et al. (2019) researches effects of lists and payoff protocols on two pairs of lotteries of Allais type and observe that, when the pair in the decision task contains a degenerate lottery, pay-one-randomly and single-task elicit different choices²¹ with or without “choice lists”

¹⁹ Their subjects were not given lottery pairs in list format; instead, the first time a subject encountered a lottery pair was on the separate screen for entering a choice from the pair.

²⁰ Subjects could not return to an earlier screen after entering their decision and advancing to the next screen.

²¹ Table 1, rows 6 and 8. list (L1) data and separate screens (S1) data are significantly different ($p=0.001$) from single-task (P1) data. Number 1 in L1, S1, P1 refers to the pair of lotteries that included a degenerate lottery: \$3 for sure vs \$4 with probability 0.8.

but the effect disappears with non-degenerate lotteries.²² They conclude that when decision tasks are embedded in lists, pay-one-randomly elicits: (i) less risk aversion in pairs with degenerate lotteries, but (ii) similar risk attitudes in pairs with non-degenerate lotteries. The latter is inconsistent with Brown and Healy's (2018) conclusion. The former is inconsistent with observed choices in Cox et al. (2015) pair 2 (\$6 vs. {\$10,0.8; \$0} and pair 5, that adds \$12 to lotteries in pair 2, (\$18 vs. {\$22,0.8; \$12}). Compared to single-task treatment, pay-one-randomly with prior information elicits significantly more risk averse behavior in pair 2 (50% vs 15% in single-task) but the effect is not significant in pair 5 (50% vs 39% in single-task). More data is needed, but it seems like the data may be telling a story of "zero effect", not "certainty effect".

5.2.4 Summary implications

We have demonstrated that pay one randomly is shown to elicit biased data in recent stress-test experiments by different teams of researchers and with varying number of decision tasks from the low of 2 in Cox et al. (2014), to 5 in Cox et al. (2015) and 7 in Freeman and Mayraz (2019), to more than 20 tasks in Brown and Healy (2018), Freeman et al. (2019) and Harrison and Swarthout (2014).

Remark 8. With or without lists, pay-one-randomly has been shown to produce biased data.

Azrieli et al. (2018) extends previous results on incentive compatibility of pay-one-randomly by requiring admissible extensions to satisfy statewise monotonicity (where states are defined by decision tasks), which is a weaker supposition than requiring that extensions satisfy the expected utility axioms. They show that if all admissible extensions satisfy statewise monotonicity then pay-one-randomly is incentive compatible (Proposition 1). Data patterns summarized in Remark 8, reject incentive compatibility of pay-one-randomly, and therefore by the contrapositive of Proposition 1, reject the supposition of statewise monotonicity of extensions.

²² Table 1, rows 7 and 9. list (L2) data, and separate screens (S2) data are not significantly different from treatment single-task (P2) data. Number 2, in L2, S2, P2 refers to the pair of lotteries with non-degenerate lotteries: \$3 with probability 0.5 vs \$4 with probability 0.4.

6. Incentive compatibility or incompatibility of other payment protocols

We next summarize what is known about incentive compatibility of several other commonly used payoff protocols.

6.1 Commonly used payment protocols

6.1.1 Single-task protocol

Each subject is asked to respond to only one choice task; outcome from a subject's chosen option is realized and is paid to the subject. Since there exists only one decision task, a subject has an incentive to reveal their latent preferences truthfully for the most preferred option available in that task. Hence the single-task protocol is incentive compatible for all decision theories, as stated in Remark 3.

6.1.2 Paying only some randomly selected subjects

Some researchers pay only *some* randomly selected subjects rather than every subject. Suppose an experimentalist has chosen well the decision tasks they want to use to research their questions of interest. And suppose each subject will be asked to make only one decision. If the researcher pays all subjects the realized outcome from their chosen option, then the experimental design provides incentives for subjects to truthfully reveal their latent risk preferences (it is incentive compatible).

Next suppose the researcher considers making one change in the above experimental design: n subjects out of the total N subjects will be randomly selected to receive salient rewards while the others will not be paid. Is the resulting experimental design still incentive compatible with addition of this subject selection protocol?

The random subject selection protocol implements the subject's chosen option with probability n/N and pays 0 with probability $1 - n/N$. Given reduction of compound lotteries, this creates prospects with a common ratio (Allais 1953). The implication for incentive compatibility depends on other features of the experimental design.

Let lotteries B and A be choice options in the experiment. With random subject selection for payoff, the choice between these two lotteries is a choice between a compound lottery, β that pays B with probability $p = n/N$ and 0 otherwise, and the compound lottery α , that pays A with probability $p = n/N$ and 0 otherwise. By expected utility axioms, if B is preferred over A , then β

is preferred over α . Hence, the protocol is incentive compatible for expected utility theory. It is *not* incentive compatible for cumulative prospect theory or rank dependent utility or dual theory of expected utility.²³

For example, let option A pay \$30 for sure and option B pay \$40 with probability 0.8. A rank dependent utility individual with power function utility, $u(x) = x^{0.6}$ and Prelec probability transformation, $f(p) = e^{-(\ln p)^{0.8}}$ will choose the safe option A in the single-task, pay-all-subjects protocol.

If only 1/3 of subjects are paid and our individual chooses A, then her payoff is \$30 if selected or \$0 if *not* selected. If she chooses B then she gets \$40 if selected and the realized payoff is \$40; she gets \$0 if selected and the realized payoff is \$0 or if she is not selected. Since $f(1/3)30^{0.6} < f(0.8/3)40^{0.6}$, our individual will choose B. Similarly, for dual expected utility, replace $u(x) = x^{0.6}$ with $u(x) = x$ in the example above and verify that A will be chosen in the single-task, pay-all-subjects protocol but B is will be chosen in the single-task experiment that pays 1/3 of subjects.

Remark 9. Randomly selecting some (but not all) subjects for salient reward payment is incentive compatible for expected utility theory but not for cumulative prospect theory or rank dependent utility or dual theory of expected utility.

Statements about incentive compatibility of payoff protocols in subsections 6.1.3 to 6.1.6 are based on proofs and counterexamples in Cox et al. (2015).

6.1.3 Pay-all-sequentially protocol

Realization of outcome from a chosen option occurs and is revealed to a subject immediately after the choice, before the next choice is made, with the sum of all realized outcomes paid to the subject at the end of the experiment session. As accumulated wealth changes between choices, this payoff protocol is not incentive compatible for expected utility of terminal wealth²⁴ (unless the subject is somehow known to have CARA preferences). Changes in wealth are not relevant to models with functionals defined on income (not terminal wealth). Hence, this protocol is incentive compatible

²³ Baltussen et al. 2012 compare pay *some* to pay *all* subjects in a single task experiment and find that it elicits less risk aversion. For a recent study that examines this protocol in a multiple price list elicitation procedure see Anderson et al. 2023. They report that it affects the ordering of subjects regarding risk aversion and that, compared to paying every subject, paying some subjects elicits less or more risk aversion depending on low or high payoff stakes.

²⁴ Classic references for the terminal wealth model are Arrow (1971) and Pratt (1964).

for expected utility of income model, commonly used in auction theory beginning with Vickrey (1961), for dual theory of expected utility (Yaari 1987), and for cumulative prospect theory with fixed reference point (“prospect theory, no memory” as in Thaler and Johnson, 1990, p. 646) but not for some other editing rules.

Remark 10. The pay-all-sequentially protocol is incentive compatible for models with functionals defined on income.

6.1.4 Pay-all-correlated protocol

Outcomes from all chosen options are realized and paid at the end of a session with one realization of the state of nature. This protocol was used for subject payoffs in experiments with comonotonic lotteries in Cox et al. (2015). They demonstrated that, with comonotonic lotteries, *pay-all-correlated* is incentive compatible for dual theory (Yaari 1987).²⁵ It is also incentive compatible for *linear* cumulative prospect theory (Schmidt and Zank 2009) if outcomes in a given state are all gains or all losses.

Remark 11. For comonotonic lotteries, the pay-all-correlated protocol is incentive compatible for dual theory and for linear cumulative prospect theory (if outcomes in a given state are all gains or all losses).²⁶

6.1.5 Pay-all-independently protocol

In many studies, outcomes from all chosen options are independently realized at the end of a session, when the sum of all outcomes is paid to the subjects. This creates portfolio incentives; *it is not incentive compatible* for any popular theory except expected value theory.

6.1.6 Pay-more-than-one randomly selected decision

This procedure creates portfolio incentives and, like pay-all-independently, is *not incentive compatible* for any popular decision theory other than expected value theory.

²⁵ With *pay-all-correlated protocol*, payoff equals the sum of payoffs (for the realized state) from all lotteries. Another way of implementation is dividing this sum of payoffs by the number of decisions; this is called PAC_N in Cox et al. (2015).

²⁶ It is not incentive compatible if there is background risk.

Remark 12. Paying all decisions independently or paying more than one randomly selected decision are incentive compatible *only for* expected value theory.

6.2 Accumulative Best Choice mechanism

As stated above, paying *one* decision *randomly* is not incentive compatible for risk preferences exhibiting nonlinearity in probabilities. Paying for *all* decisions *sequentially* is not incentive compatible for risk preferences nonlinear in accumulated monetary payoffs. One might envision paying for only one decision task but using some non-random selection procedure that preserves incentivization of each decision task. Li (2021) presents such a mechanism, the Accumulative Best Choice mechanism.

Li (2021) shows that Accumulative Best Choice is incentive compatible for complete and transitive preferences. Thus, it can be used to elicit risk preferences, whether they are nonlinear in probabilities, payoffs, or both. How does it work? After any decision task, the feasible set for the next task includes the chosen option. Accumulative Best Choice pays only the last choice (that is, one task) while preserving incentives for every task. There are, however, limitations on experimental designs where this protocol can be applied. The endogenous formation of decision sets may not necessarily be consistent with some research questions.²⁷

Li (2021) reports two experiments with Accumulative Best Choice as a payoff protocol. In one experiment she uses Cox et al. (2015) five lottery pairs whereas in the other experiment the lotteries are from Harrison and Swarthout (2014). In both experiments, she finds that risk preferences elicited with Accumulative Best Choice and single-task incentives are *not* statistically different. This finding complements the Cox et al. (2015) and Harrison and Swarthout (2014) finding that pay-one-randomly does elicit different risk preferences than the single-task protocol.²⁸

Remark 13. The accumulative best choice mechanism is incentive compatible for all choice theories with first-order stochastic dominance.

²⁷ Li (2021) uses Accumulative Best Choice (ABC) to elicit choices from pairs of lotteries in Table 1 of Cox, et al. (2015) that include Allais (1953) type of lotteries. It is necessary to elicit choices across pairs of lottery pairs to test for Allais-paradoxical responses. This was possible with ABC for the Cox, et al. lotteries only because the lottery pairs in their table exhibit first-order stochastic dominance ordering across rows that could be used in constructing the sequence of lottery pairs in the experiment.

²⁸ This result holds for the Cox et al. (2015) treatment in which subjects were *not* given prior information about lottery pairs.

7. Why is the incentive compatibility issue important?

The mounting evidence of risk preference elicitation bias from random task selection has stimulated research on identifying new payoff mechanisms that are incentive compatible for non-EU decision theories and on experiments with their behavioral properties. Some progress has been made. Cox, et al. (2015) introduce two versions of a payoff mechanism that is (weakly) incentive compatible for models that satisfy dual independence (Yaari 1987). More importantly, Li (2021) introduces the accumulated best choice mechanism that is incentive compatible for all decision theories that have complete and transitive preferences and can be used to elicit risk preferences in many types of experiments.

It has been about forty years since Holt (1986) stated why conclusions that can be correctly drawn from data can depend on the subject payment protocol used in an experiment. In following decades there has been additional theoretical work on payment protocols and data from experiments on payoff protocol effects on behavior. To date, this literature has had only modest effect on practice in experimental economics.

A survey in Azrieli, et al. (2018, pgs. 1473-1474) of papers on decisions under risk published in 2011 in top five general readership journals and *Experimental Economics* reveals use of many distinct payoff protocols. In about 29 percent of papers in the survey, authors did not describe how subjects were paid in many-decision experiments. Another 48 percent of papers described how subjects were paid but provided no justification for the chosen mechanism. In those papers with identifiable payoff mechanisms: 56 percent paid for every decision, 25 percent paid for one randomly selected decision, and 13 percent paid for more than one randomly selected decision. The high incidence of unexplained and divergent choices of payoff protocols suggests researchers may consider this to be an “unimportant detail” of experimental methods. In more recent years, have researchers used incentive compatible payoff protocols?

Top general interest journals are continuing to publish papers that draw conclusions with unknown validity because of use of incentive incompatible payoff protocols. An example of how such papers pass through the review process even when the problem is acknowledged is provided by Bernheim and Sprenger (2020). They use pay-one-randomly in an experiment on rank-dependent probability weighting. In what one can guess is a response to reviewer comments, the

authors (pg. 1380) acknowledge the incentive incompatibility issue.²⁹ But the authors, reviewers and editor proceed to publish the paper, and draw conclusions from the data, without requiring an additional experiment treatment to assess the empirical significance of biased incentives. The authors could have implemented an approach similar to Harrison and Swarthout (2014) to assess and correct for significant bias in their data by implementing a single-task control treatment.

We recently surveyed papers published in *Experimental Economics* during 2016 – 2024. Here is what we found. On a positive note: (1) the payment protocol used in an experiment was usually clear from the text or the subject instructions or both; and (2) there were many papers with models based on expected value theory for which all commonly used payoff protocols are incentive compatible. Choice from among alternative payment protocols varied a lot across papers. There were many heuristic experiments designed to collect interesting behavior but not to test or apply formal models. For these papers, theory is not a guide for choice of payoff protocol but empirical properties of payoff protocols reported in our cited literature (in section 5) is informative. On a disappointing note, a substantial number of papers (38) used incentive-*incompatible* payoff protocols, almost half of them by using *pay one randomly* for non-expected utility models (16 papers) and the others distributed between *paying two or more randomly* for expected utility models, *paying two or more randomly* for rank dependent utility models, *paying all independently* at the end, and combination of *paying one randomly* and *paying all independently* at the end, and *paying all sequentially* for expected utility of terminal wealth models.³⁰

There was virtually no discussion in any of the surveyed papers of reasons why a particular payoff protocol was chosen, nor any attempt to test for potential biases in the data. It appears many researchers may believe that one's choice from among familiar payment protocols is not relevant to validity of tests of hypotheses with experimental data. Except for testing expected value models, such belief is demonstrably incorrect.

8. Suggestions for experimental methods

Incentive compatibility or incentive incompatibility of payoff protocols follows from theory. Whether use of an incentive incompatible payoff protocol yields data that is *significantly* biased is

²⁹ The authors cite Starmer and Sugden (1991) and Cubitt et al. (1998) as demonstrating that subjects isolate each decision; in fact, the former paper states (pg. 977) that it *cannot be* concluded from their data that the random-lottery incentive system is unbiased. More tellingly, the authors do not cite any of the many papers that conclude subjects do *not* isolate decisions: see section 5.1.2 above for such citations.

³⁰ We did not count the number of papers paying only some but not all of the subjects.

a distinct, empirical question.

8.1 Avoiding incentive incompatible protocols

A straightforward recommendation is that researchers avoid using incentive incompatible protocols when possible. Randomly selecting *two or more decisions* for payoffs introduces portfolio effects into the incentives in the experiment. This is easy to avoid by randomly selecting *only one* decision when testing EU or using it for modeling. For cumulative prospect theory or rank dependent utility use the Accumulated Best Choice mechanism or single-task, *which are incentive compatible*, not the pay one randomly protocol which is not.

Paying *all decisions at the end* with independent random draws creates portfolio incentives. This should be avoided unless one is testing a theory of portfolio choice.

For income models, use pay *all decisions sequentially*. It is incentive compatible for the EU income model (as in most auction market theory), for dual theory (non-linear probabilities) or cumulative prospect theory with fixed reference point but not other editing rules. For dual theory, an alternative incentive compatible payoff protocol (for comonotonic lotteries) is pay *all decisions correlated*.

Randomly selecting (*only*) *some subjects* for payoff creates prospects of Allais (1953) type and brings EU “in the back door”. This should be avoided in (single- and many-tasks) experiments on non-EU models. There is a simple, although more expensive, solution: pay all the subjects.

8.2 Checking for significance of bias from incentive incompatible protocols

Bias from using an incentive incompatible payoff protocol may or may not be economically or statistically significant. There are known methods for checking that can be incorporated into experimental designs. The preference reversal literature provides an example. Cox and Epstein (1989) and Cox and Grether (1996) needed to avoid random decision selection to distinguish between response mode effects and inconsistencies with EU for reasons explained by Holt (1986). They used pay-all-sequentially, which could introduce bias from income effects. They analyzed the data for income effects, and for order effects, which were found to be insignificant. Hence the sequential choices with different response modes could be interpreted as response mode effects. Harrison and Swarthout (2014) used single-task as a control treatment in a risk attitude experiment (otherwise) using random decision selection. The single-task data made it possible to measure the significance of bias in econometric estimation of the RDU model parameters using random

decision selection data. Other papers that have paid a single known decision out of several as a control treatment to ascertain how payment protocols affect decisions include Beattie and Loomes (1997), Cubitt et al. (1998), and Cox et al. (2015). With single-task, a subject makes a single decision whereas payment protocols are used in within-subjects designs that require subjects to make many decisions. In a variant of pay-one-known, which we have called Impure OT (Cox et al. 2015), a subject makes several hypothetical decisions followed by a single decision with salient rewards. It has been used as a control treatment in Starmer and Sugden (1991), Cubitt et al. (1998), Cox et al. (2015), Brown and Healy (2018) and Freeman and Mayraz (2019).

9 Concluding remarks

We present a finding in the literature in which data from paying salient rewards clearly leads to the opposite conclusion than does use of hypothetical payoffs about a central research question – and we describe the features of the experimental task that produced this outcome. This finding is reported in the literature on the preference reversal phenomenon, the literature that initiated inquiry into how the validity of conclusions that can be drawn from experimental data depends on the subject payment protocol used in the experiment. We take up this topic in an extension and critical discussion of theoretical and behavioral properties of subject payment protocols. As we explain, which payment protocols are incentive compatible – and which ones incentivize biased responses – depends on the research question and relevant theoretical models. There is no unique payoff protocol that is appropriate for all research questions. It follows that choice of payoff protocol should be considered an essential feature of experimental design. Surveys of recent literature suggests it often is not considered important.

References

- Agranov, M. and Ortoleva, P., 2017. Stochastic choice and preferences for randomization, *Journal of Political Economy* 125, 40-68.
- Agranov, M. and Ortoleva, P., 2022. Revealed preferences for randomization: An overview. *AEA Papers and Proceedings* 112, 426–30.
- Anderson, L.R., Freeborn, B.A., McAlvanah, P. and Turscak, A., 2023. Pay every subject or pay only some? *J Risk Uncertain* 66, 161–188.
- 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.
- Arrow, K., 1971. *Essays in the Theory of Risk-Bearing*. Markham, Chicago.
- Azrieli, Y., Chambers, C.P. and Healy, P.J., 2018. Incentives in experiments: a theoretical analysis. *Journal of Political Economy* 126 (4), 1472-1503.
- Azrieli, Y., Chambers, C.P. and Healy, P.J., 2020. Incentives in experiments with objective lotteries. *Experimental Economics* 23, 1–29.
- Baltussen, G., Post, G.T., van den Assem, M.J. and Wakker, P.P., 2012. Random incentive systems in a dynamic choice experiment. *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, 155-168.
- Becker, G.M., DeGroot, M.H., and Marschak, J., 1964. Measuring utility by a single-response sequential method". *Behavioral Science* 9 (3): 226–32.
- Bernheim, B.D and Sprenger, C.S., 2020. On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting. *Econometrica*, 88(4), 1363-1409.
- Brañas-Garza, P., Jorrat, D., Espín, A.M. and Sánchez, A., 2023. Paid and hypothetical time preferences are the same: lab, field and online evidence. *Experimental Economics* 26, 412–434.
- Brown, A.L. and Healy, P.J., 2018. Separated decisions. *European Economic Review* 101(c), 20-34.
- Charness, G., Eckel, C., Gneezy, U. and Kajackaite, A. 2018. Complexity in risk elicitation may affect the conclusions: A demonstration using gender differences. *Journal of Risk and Uncertainty* 56, 1–17.

- 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.M., 1996. The Preference Reversal Phenomenon: Response Mode, Markets and Incentives. *Economic Theory* 7 (3):381-405.
- Cox, J.C. and Sadiraj, V., 2019. Incentives. Chapter 1 in *Handbook of Research Methods and Applications in Experimental Economics*, Arthur Schram and Aljaž Ule (eds.), Edward Elgar Publishing.
- Cox, J.C., Sadiraj, V. and Schmidt, U., 2014. Asymmetrically dominated choice problems, the isolation hypothesis and random incentive mechanisms. *PLoS ONE* 9(3): e90742.
- Cox, J.C., Sadiraj, V. and Schmidt, U., 2015. Paradoxes and mechanisms for choice under risk. *Experimental Economics* 18(2), 215-250
- Cubitt, R. P., Starmer, C. and Sugden, R., 1998. On the validity of the random lottery incentive system. *Experimental Economics* 1, 115–131.
- Freeman, D.J., Halevy, Y. and Kneeland, T., 2019. Eliciting risk preferences using choice lists. *Quantitative Economics* 10(1), 217-237
- Freeman, D.J. and Mayraz, G., 2019. Why choice lists increase risk taking. *Experimental Economics* 22(1), 131-154
- Grether, D.M. and Plott, C.R, 1979. Economic Theory of Choice and the Preference Reversal Phenomenon. *American Economic Review* 69 (4), 623-38.
- Hackethal, A., Kirchler, M., Laudenbach, C., Razen M., and A. Weber., 2023. On the role of monetary incentives in risk preference elicitation experiments. *Journal of Risk and Uncertainty* 66, 189–213 (2023).
- Harrison, G.W. and Swarthout, J.T., 2014. Experimental payment protocols and the bipolar behaviorist. *Theory and Decision*. 77(3), 423-438.
- Hey, J.D. and Lee, J., 2005a. Do subjects separate (or are they sophisticated?) *Experimental Economics* 8(3), 233-265.
- Hey, J.D. and Lee, J., 2005b. Do subjects remember the past? *Applied Economics* 37, 9-18.
- Holt, C.A., 1986. Preference reversals and the independence axiom. *American Economic Review*. 76(3), 508-515.
- Holt, C.A. and Laury, S.K., 2002. Risk aversion and incentive effects. *American Economic Review* 92, 1644–1655.

- Karni, E. and Safra, Z., 1987. "Preference reversal" and the observability of preferences by experimental methods. *Econometrica* 55(3), 675-685.
- Li, Y., 2021. The ABC mechanism: an incentive compatible payoff mechanism for elicitation of outcome and probability transformations. *Experimental Economics* 24, 1019-1046.
- Lichtenstein, S. and Slovic, P., 1971. Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89, 46-55.
- Machina, M.J., 1985. Stochastic Choice Functions Generated From Deterministic Preferences Over Lotteries." *Economic Journal* 95 (379), 575-94.
- Pratt, J.W., 1964. Risk aversion in the small and in the large. *Econometrica* 32, 122-136.
- Quiggin, J., 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3(4), 323-43.
- Schmidt, U., and Zank, H., 2009. A Simple Model of Cumulative Prospect Theory. *Journal of Mathematical Economics* 45, 308-319.
- Seidl, C., 2002. Preference Reversal. *Journal of Economic Surveys*, 16: 621-655.
- Slovic, P. and Lichtenstein, S., 1968. Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology Monograph*, 78, 1-18.
- Smith, V.L., 1982. Microeconomic Systems as an Experimental Science. *American Economic Review* 72 (5): 923-55.
- Starmer, C. and Sugden, R., 1991. Does the random lottery incentive system elicit true preferences? An experimental investigation. *American Economic Review*. 81(4), 971- 978.
- Sugden, R., 2015. Looking for a Psychology for the Inner Rational Agent. *Social Theory and Practice*, 41(4), 579-598.
- Thaler. R.D. and Johnson, E.J., 1990. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science* 36, 643-660.
- Tversky, A. and Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297-323.
- Tversky, A., Slovic, P. and Kahneman, D., 1990. The causes of preference reversal. *American Economic Review* 80, 204-17
- Vickrey, W., 1961. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance* 16, 8-37.
- Yaari, M.E., 1987. The dual theory of choice under risk. *Econometrica* 55(1), 95-115.

Table 1. Recent Experimental Tests for Pay-One-Randomly Effect on Choices

Paper	Row	Treatment Comparison	Certainty Effect	List Display	Separate Choice Screens	Menu Effects	p-value
Brown, Healy (2018)	1	L-RPS vs. L-14	no	yes, yes	no, no	yes, yes	0.041
	2	S-RPS vs. S-14	no	no, no	yes, yes	yes, yes	0.697
Cox, et al. (2015)	3	PORpi vs. ImpureOT	no	yes ^a , yes ^a	yes, yes	yes, yes	0.643
	4	PORnp vs. OT	no	no, no	yes, yes	no ^b , no	0.001
	5	1 st Choice PORnp vs. OT	no	no, no	yes, yes	no, no	0.001
Freeman, et al. (2019)	6	L1 vs. P1	yes	yes, no	no, yes	yes, no	0.001
	7	L2 vs. P2	no	yes, no	no, yes	yes, no	0.17
	8	S1 vs. P1	yes	yes, no	yes/no ^c , yes	yes, no	0.001
	9	S2 vs. P2	no	yes, no	yes/no ^c , yes	yes, no	0.21
Freeman, Mayraz (2019)	10	Q85: R-list vs. SC	yes	yes, no	no, yes	yes, no	0.001
	11	Q85: K-list vs. SC	yes	yes, no	no, yes	yes, no	0.013
Harrison, Swarthout (2014)	12	1 in 30 vs. 1 in 1	no	no, no	yes, yes	yes, no	0.055

^aNote that subjects were not presented with a list, but they could construct lists from separate slips of paper for individual lottery pairs if they wanted to.

^bSubjects' memory of previous choice pairs could have a "menu-type" effect. "Decisions first made on separate screens then reported in a filled-in choice list with opportunity to change choices.

Acronyms: Display of decision tasks: S Separated; L list.

Pay-one-randomly: 1 in 30, RPS, R-list, POR, PORnp (no prior information), PORpi (subjects saw all five tasks before making decisions)

More than one decision but only one pays and subject knows which one: K-list (Q85), L1, L2, L-14, S-14, ImpureOT

Single-task treatments: SC, OT, P1, P2, 1 in 1