1. Incentives*

James C. Cox and Vjollca Sadiraj

1 INTRODUCTION

Incentives are a major part of what defines the "economics" part in experimental economics. A "salient" incentive payment varies with the outcomes in an experiment that result from actions by the subjects (Wilde, 1981).[1] A participation fee that does not vary with subjects' actions is not salient. Economic experiments commonly, but not always, use monetary incentives.[2] Since salient monetary incentives vary with outcomes that result from subjects' actions, they provide motivation to subjects not satiated in money (Smith, 1976) for undertaking actions within an experiment.

Smith (1976) identifies three complications (or "qualifications") that can confound control in an experiment: (1) subjective cost; (2) interpersonal utilities; and (3) value of playing the game.[3] Fouraker and Siegel (1963) and Smith (1976) stress the importance of (4) complete vs incomplete information. We add (5) response mode effects (Grether and Plott, 1979; Cox and Grether, 1996) and (6) game-form misconception (Plott and Zeiler, 2005; Cason and Plott, 2014).

In the following sections, we provide examples of how such complications interact with experimental methods in several topic areas of research. The discussion will focus on experimental approaches to control for confounds. The choice of the salient payoff level (dominance of the reward; Wilde, 1981) is used as an important instrument to offset subjective costs. Other instruments include: withholding information on others' payoffs (privacy; Wilde, 1981) to inhibit triggering interpersonal preferences (when they are not being studied); practice rounds or repetition to promote understanding; alternative institutional formats (e.g., game trees and price clocks) to vary accessibility of payoff information; and binary lottery payoffs to incentivize risk neutrality (when it is being assumed). We discuss how salient payoffs can interact with experimental task repetition and economic institutions to affect subjects' behavior. The last section discusses problems with the common practice of experimentalists using payoff protocols in experiments with

* Andreas Ortmann, an anonymous reviewer, and an editor provided helpful comments and suggestions on a preliminary draft of this chapter. We thank the National Science Foundation (grant number SES-1658743) for research funding.

1 Saliency, dominance, and privacy are among the precepts discussed by Smith (1982a), who attributes these three to Louis Wilde. For scholarly accuracy we follow Smith and cite Wilde (1981). (The correct year of publication and edited book title for Wilde's seminal chapter are included below in our References.)

2 Paying money to subjects can, potentially, lead to self-selection by participants more strongly motivated by financial compensation. Such possible selection may or may not be a problem depending on the research question addressed in an experiment. One topic area in which this type of subject self-selection could be a concern is in experiments on eliciting social preferences. Falk, Meier and Zehnder (2014) report little selection effect in a social preferences experiment.

3 See also Siegel (1961) and Fouraker and Siegel (1963). In today's terminology, "interpersonal utilities" would be labeled other-regarding preferences or social preferences.

more than one decision that are not incentive compatible for eliciting true preferences within each decision.

In discussing incentives, it is natural to view the experimenter as a principal and the subjects as agents. The experimenter controls the set of available actions and the mapping from actions to outcomes, and wants the subject to choose his or her most preferred action assuming he or she completely understands the environment and institution and has zero subjective cost from experimental task completion. We call this the "experimental decision problem."

One difficulty is that the experimental tasks given to subjects may be unfamiliar to them, and it may require effort to understand the tasks and the implications of alternative actions. If the subject incurs a subjective cost from experimental task completion then his or her choice will be the most preferred, given his or her subjective effort cost, which may confound interpretation of the data because such cost is not observable.[4] We call this the "subject's decision problem." In this way the existence of subjective cost creates a moral hazard problem the experimenter needs to address. One way to better align the incentives of the subjects and the experimenter is through dominant salient rewards that compensate for subjective cost by increasing the opportunity cost of actions that are suboptimal for the experimental decision problem.

If a lazy or distracted subject does not understand the experimental environment or institution then providing larger payoffs may motivate effort leading to better understanding.[5] In addition, scaled-up payoffs can reduce decision errors from random choices among actions with trivial economic consequences. In contrast, larger payoffs may not improve quality of data if a subject believes he or she understands the experimental environment and institution but is mistaken. If there is "failure of game-form recognition" (Cason and Plott, 2014) then scaled-up payoffs may not be helpful, but providing subjects with opportunities for learning may produce data that more accurately reflect the subject's preferences.[6] This issue is discussed below (in Section 8). Paying subjects real salient payoffs, by itself, does not correct for response mode effects such as preference reversals (Grether and Plott, 1979).[7] In contrast, the preference reversal phenomenon is not robust to repetitive decisions in markets with large salient payoffs (Cox and Grether, 1996).[8]

In many-decision experiments, a subject who does understand the experimental environment and institution will be motivated to provide biased data (that is, make choices that do not reflect his or her true preferences at the single decision level) if the experimenter uses a

---

4 Smith (1976); see also Camerer and Hogarth (1999).

5 See Ariely et al. (2009) for discussion of the boundaries of the effectiveness of this approach.

6 There is some evidence that eliciting beliefs about play by others can elicit choices more consistent with theoretical predictions (e.g., Croson, 2000 finds that it promotes defection and free riding) but eliciting beliefs can cause cross-task contamination from hedging. Space limitations prevent us from discussing these issues. For a survey on belief elicitation, see Schlag, Tremewan and Van der Weele (2015).

7 A preference reversal occurs when a subject chooses lottery *a* over lottery *b* but places a higher selling price on the latter.

8 Evidence on the effect of salient payoffs on choices is mixed (Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001). Camerer and Hogarth (1999) cover 74 studies: 23 studies report a positive effect of incentives on mean performance, 27 studies find no effect, and nine studies report negative effects; the remaining 15 studies are categorized as "having effect but no performance standard." Hertwig and Ortmann (2003) find that positive effects are mainly observed in "judgments and decisions" studies, whereas evidence on "no effects" comes from studies of "games and markets." Smith and Walker (1993) report that larger monetary incentives decrease variance of responses.

mapping from multiple choices to payoffs – a payoff protocol or payoff mechanism – that is *not* incentive compatible.[9] Use of incentive-*in*compatible payoff protocols is widespread in the literature, most especially in experiments on decisions under risk, but also in many other topic areas (Cox, Sadiraj and Schmidt, 2015). This issue is discussed in Section 9.

We next present an historically important example that illustrates use of induced valuation to create controlled supply and demand in markets. The experimenter's objective is to learn about the efficiency properties of market institutions. The induced supply and demand schedules make efficiency of market allocations unambiguously measurable by researchers.

2 INDUCED VALUATION IN LABORATORY MARKET EXPERIMENTS

In a laboratory market experiment a researcher assigns values to a buyer for each unit of an abstract commodity that the buyer can buy, V_j, for $j = 1, 2, \ldots n$:

$$V_1 \geq V_2 \geq V_3 \geq \ldots \geq V_n \tag{1.1}$$

Similarly, costs are assigned to a seller for each unit that the seller can sell, C_i, for $i = 1, 2, \ldots m$:

$$C_1 \leq C_2 \leq C_3 \leq \ldots \leq C_m \tag{1.2}$$

This provides an example of the experimental method called "induced valuation" (Smith, 1976). These values and costs are ordinarily the private information of buyers and sellers.

Salient incentives for exchange are created by an experimenter's credible promise to pay subjects as follows. If a buyer with value V_k exchanges a single unit with a seller with cost C_j at the price P then the experimenter pays them the monetary amounts:

$$\begin{aligned} &Buyer\text{: } V_k - P \\ &Seller\text{: } P - C_j \end{aligned} \tag{1.3}$$

So long as the subjects prefer more money to less, and transaction costs are negligible, the experimenter's credible promise to pay subjects the monetary amounts given by (1.3) provides incentives to buyers and sellers to search for better deals and induces on buyers and sellers the demand and supply prices given by statements (1.1) and (1.2).

Assume a researcher hopes to learn about the efficiency properties of a market institution by conducting an experiment with controlled demand and supply functions created with induced valuation.[10] One important market institution that reduces transaction and

[9] In the case of a multi-decision experiment, we say that a payoff protocol is *not* incentive compatible if it provides incentive for a subject with true preference for option *a* over option *b* to choose (. . . ,*b*, . . .) over (. . . ,*a*, . . .) when that choice is embedded as one of many and payoffs are generated with the protocol.

[10] The observable maximum possible gain from exchange G_{max} is given by the area between the market demand and supply functions created by induced valuation. The observable realized gain from exchange G_{obs} is given by the total amount of money paid to the experimental subjects following the rules given by statements (1.3). The measure of efficiency of exchange is $E = 100 \times G_{obs} / G_{max}$.

search costs by centralizing information on bids and asks is the double auction used on most stock exchanges.[11] A robust finding is that laboratory double auction allocations (of abstract commodities) are highly efficient, with efficiencies close to 100 percent within a few repetitions (Smith, 1982a, 1982b).

A stress test of the robustness of high efficiency of the double auction is provided by a boundary experimental design in which all values are equal to V and all costs are equal to C, with $V > C$, but there are more units on one side of the market (Smith, 1976, 1982a). With this design, at competitive equilibrium all gains from exchange go to the short side of the market. With privately induced demand and supply schedules, high efficiencies in double auction markets are robust to this boundary experiment, at least when small "commissions" are paid to overcome transaction costs. Making the values and costs public information, however, impedes convergence in this boundary experiment, which seems to reflect the effects of interpersonal utilities. This illustrates one reason experimental researchers usually implement induced values and costs as private information in order to inhibit possible effects of interpersonal utilities on exchange.[12]

3 INTERPERSONAL UTILITIES AND INFORMATION

3.1 Duopoly Games

Fouraker and Siegel (1963) conducted several experiments on duopoly (and triopoly) designed to study the effect of information about others' payoffs on decisions. They argue that predictions of models that assume simple own payoff-maximizing agents are more likely to be observed in the absence of information on payoffs of other players. Complete information about payoffs offers opportunities for interpersonal comparisons, which leads to higher dispersion of outcomes as more instances of cooperative or competitive strategies are elicited. Fouraker and Siegel identify three types of subjects (or strategies): simple maximizers (own payoff), rivalrous competitors (own minus others' payoff), and cooperators (own plus others' payoff).

To investigate whether interpersonal preferences were the source of the diversity of the data in the complete information treatment, Fouraker and Siegel used incentives (in the form of bonuses) to induce cooperative or rivalrous behavior. To induce cooperation, any pair of subjects with total payoff exceeding a certain level of profit received an additional bonus. To induce rivalry, within each pair of subjects the subject with the larger profit received a bonus. Subjects participated in ten practice rounds with feedback before they played one real monetary payoff round. Their data provide support for the validity of:

---

[11] The permissible actions, pricing and contracting rules for a simplified version of the double auction institution in which each exchange is a single unit (but buyers and sellers can make multiple exchanges in a market period) are as follows (Smith, 1982a). (1) Any buyer can: (a) make a bid to buy a unit at a stated price or (b) accept an outstanding offer price, which makes a market price, and effects an exchange of a unit. (2). Any seller can: (a) make an offer to sell a unit at a stated price or (b) accept an outstanding bid price, which makes a market price and effects an exchange of a unit.

[12] Outside the laboratory, the producer or consumer surplus realized in a market exchange is ordinarily private information.

(1) Cournot solution in the incomplete information treatment; (2) cooperative solution in induced cooperation treatment; and (3) competitive solution in induced rivalry treatment. An alternative approach, often used these days, is to first classify subjects according to their types and then use that information to predict (or explain) behavior in the game of interest. This approach was rejected by Fouraker and Siegel on two grounds: (1) "a priori preference for the experimental rather than the psychometric approach" (Fouraker and Siegel, 1963, p. 155), and (2) absence of a reliable method for type classification of subjects.

3.2 Centipede Games

Several studies of play in centipede games[13] offer examples of use of the experimental approach to address issues with interpersonal utilities and subjective costs. Suppose the research question is identifying conditions that accelerate breakdown of a joint enterprise that can be modeled as a centipede game. One might expect the play of the game to be affected by efficiency (e.g., exponentially increasing centipede games), when total payoffs increase as the game continues, as well as by asymmetry of the distribution of the total payoff (e.g., competitive centipede games).

Krockow, Colman and Pulford (2016) provide a survey of experimental studies of centipede games from 1992 to 2016 covering 72 treatments. Play in competitive centipede game experiments reported by Fey, McKelvey and Palfrey (1996) and Cox and James (2012) is closest to complete unraveling to a take at the first opportunity.

To control for efficiency considerations, Fey et al. (1996) study play in a constant-sum centipede game, in which "not quitting" never increases total payoff to all players. The unique outcome of Nash equilibrium (with selfish, fair or efficiency preferences) of this game is quitting at the very first node. They find that in the six-node and ten-node games, after the fifth match the average exit rates at the initial node are 0.70 and 0.53.[14] In comparison, the exit rates in the first five matches are 0.46 (six-node game) and 0.37 (ten-node game). These figures highlight the importance of providing subjects with opportunities for learning about the structure of the game and the play of others by repeatedly playing the game.

The experiment reported by Cox and James (2012) preserves efficiency but: (1) induces rivalry by allocating all gains to the quitter; (2) discourages "never quit" with zero end-node payoffs (Aumann, 1992); and (3) uses private payoff information. Two institutional formats, tree and clock, are used to study play of their exponential, zero end-node payoff centipede game (see Figures 1–3 in Cox and James, 2012).[15] The clock format needs more cognitive effort to appreciate efficiencies lost by quitting at early nodes as it requires a mental picture of the future payoffs; such payoffs are visually explicit in the tree. Within a couple of matches, complete unraveling of play is observed in simultaneous move games as well as in sequential move games with clock format (see Figures 6 and 7 in Cox and

13 The game was introduced by Rosenthal (1981) to challenge common knowledge of rationality.

14 The maximum payoff in the game is $2.92. Subjects participated in nine to ten matches with randomly matched opponents.

15 The tree format is an extensive-form game tree that provides payoff information for all decision nodes in the game. The clock format, conventionally used in Dutch auctions, provides payoff information only for the currently active decision node.

James, 2012).[16] A slower convergence to complete unraveling is observed in the tree format with simultaneous moves.

Two studies that use payoff level as an instrument in their experimental approach are Rapoport et al. (2003) and Palacios-Huerta and Volij (2009). Rapoport et al. (2003) conduct three-player, zero end-node payoff centipede games with exponentially increasing payoffs, vary the scale of payoffs, and provide abundant opportunities for behavior to stabilize.[17] In the high payoff experiment, in the last 15 matches, 72 percent of exits were observed at the initial node. In the regular payoff experiment only 2.6 percent of exits were at node 1. They conclude that increasing (1) payoff level, (2) number of players, and (3) number of matches accelerates unraveling. Palacios-Huerta and Volij (2009) use both experienced (chess player) and regular (student) subjects in play of an exponentially increasing centipede game with payoffs ten times the ones used in the first study of centipede game by McKelvey and Palfrey (1992). When chess masters play each other in the field experiment with only one match, more than 73 percent of the exits are at the initial node. In the laboratory, after the fifth match, when chess masters play each other the exit rate at the initial node is 100 percent but down to 60 percent when chess masters play students.[18] Despite the high payoff levels, data from students playing students are similar to the ones observed by McKelvey and Palfrey.[19] Palacios-Huerta and Volij argue against social preferences and favor own rationality and belief in others' rationality as accounting for play in perfect information games.

4 INTERPERSONAL UTILITIES AND EMPIRICAL VALIDITY OF SOLUTION CONCEPTS

The standard ultimatum bargaining game (UBG) models a one-time interaction between two agents: one proposing an allocation $((1-r)S, rS)$ of a certain amount of money, S and the other given the power to veto implementation of the proposal. The standard dictator game (DG) takes the veto power away from the non-proposer. Because there are no strategic considerations in the DG, the game has been a favorite tool in measuring interpersonal utilities.

Nash equilibrium is of little help for modeling play in UBG because it can support any outcome $((1-r)S, rS)$, but subgame perfect equilibrium (Selten, 1965) makes sharper predictions.[20] Starting with Güth, Schmittberger and Schwarze (1982), play in UBG turned out to be a difficult problem for induced valuation in the laboratory, setting the stage for an extensive research program pushing the idea that interpersonal utilities should be incorporated into players' utilities of outcomes to make proper inferences from the

16 The maximum payoff in the game is $20. Subjects participated in ten matches after three practice rounds.

17 The maximum payoffs used in the high-stakes and regular-stakes treatments in Rapoport et al. (2003) were $2560 and $25.6. Subjects participated in 60 matches.

18 Caution is needed in interpreting this result because Levitt, List and Sadoff (2011) did *not* find that play by chess masters unraveled in their centipede game experiment. They did find that many chess masters backward induct in their race to 100 game.

19 See also Smith (2010, pp. 5–8) for a discussion on backward induction and its empirical validity being regularly challenged by data.

20 Subgame perfect equilibrium is theoretically one of the most embraced solution concepts of games with perfect information.

data. As discussed above, the early approach (Smith, 1976) to control for interdependencies was through privacy (i.e., no information on others' payoffs). The usefulness of this approach, however, is limited when the research question is empirical validity of a solution concept which usually requires knowledge of payoffs of others.

There have been numerous studies of play in UBG with differing payoff scales.[21] In a survey of 37 studies of UBG, Oosterbeek et al. (2004) find that: (1) shares are around 40 percent and decrease in amount available to share (S); and (2) rejection rates are about 16 percent. In a recent study motivated by reports of insensitivity of responder behavior to payoff scale,[22] Andersen et al. (2011) attribute the insensitivity to scarcity of offers below 20 percent. They find support for the use of very large payoff levels on decreasing rejections.[23] Unfortunately, as they do not have a dictator game to compare distributions of offers in DG and UBG it remains open whether their data from the treatment with the largest payoff scale support subgame perfect equilibrium.

Rather than interpersonal utilities, Hoffman et al. (1994) explore several hypotheses about effects of experimental protocols on play. The instruments used by Hoffman et al. to control for confounds include: neutral wording, exchange framing, contest role assignment, and double-blind payoff protocol to control for experimenter "audience" effect.[24] Lower offers and higher acceptance rates were observed with contest entitlements, double-blind payoffs and exchange framing.[25]

In the same spirit, Stahl and Haruvy (2008) argue that behavior in UBG reflects unintended induced social context by the verbal description of the game. They implemented UBG with a game tree representation in addition to a standard verbal description of the game. Observed play was similar to other studies of UBG in the verbal representation but close to subgame perfect equilibrium with selfish preferences in the tree representation.[26] Stahl and Haruvy used both conventional bargaining over shares of the pie as well as bargaining over probabilities of receiving the whole pie (in binary lotteries). They find no effect on proposers or responders from this use of binary lottery payoffs. Binary lotteries have been proposed as an instrument for inducing risk neutrality, an issue that we take up in Section 6.

While many researchers (Hoffman et al., 1994; Stahl and Haruvy, 2008; Smith, 2010) take the position that what appears as "interpersonal utility" may reflect arbitrary role

---

21 For surveys, see Oosterbeek, Sloof and Van de Kuilen (2004) and Karagözoglu and Urhan (2017); for a meta-study of UBG see Tisserand (2016).

22 Cited examples include Slonim and Roth (1998), Cameron (1999) and Munier and Zaharia (2002).

23 Average (aggregated) rejection rates across their four payoff levels were: 36 percent, 43 percent, 28 percent and 4 percent. The dollar equivalents of the rupee payoffs {20, 200, 2000, 20 000} were {\$0.41, \$4.1, \$41, \$410}. Daily average wage in India was 100 rupees.

24 In a double anonymous (or "double-blind") payoff protocol subjects' actions are anonymous to everyone, including the experimenter. It is possible to pay the subjects salient rewards from an experiment but preserve their anonymity by the following procedure. Have each subject choose one from among N identical-looking sealed envelopes. Each envelope contains a uniquely numbered key. Subjects are instructed to keep their key numbers private but enter them in their computers or paper response forms. At the end of the experiment, a subject uses the key to open, in private, a numbered mailbox containing salient rewards from their participation in the experiment.

25 In \$10 games, the percentage of offers greater than or equal to \$4 were 45 percent (contest and exchange) and 85 percent (random role assignment and "divide \$10" wording).

26 In the tree representation 69 percent of proposers chose the most selfish available offer and 96 percent of responders accepted the offer. In the verbal representation, these figures were down to 23 percent and 49 percent.

assignments, social norms triggered by the presence of audience (experimenter) effects, or suggestive description of the decision task, others (Kahneman, Knetsch and Thaler, 1986; Bolton, 1991; Rabin, 1993) have shifted towards inclusion of interpersonal utilities in the experimental decision problem. This initiated an explosive research program on assessing interpersonal utilities via experiments with the dictator game (DG).[27]

The generosity of subjects' choices, however, can depend on the protocol used to assign endowments in an experiment. This has been clearly established in experiments on DG that compare data from treatments in which subjects earn their endowments rather than having them randomly assigned by the experimenter. Cherry, Frykblom and Shogren (2002) had subjects who would subsequently make decisions in a dictator game, first earn their endowments by answering questions from the Graduate Management Admissions Test (GMAT). Oxoby and Spraggon (2008), in addition, had recipients earn money by answering GMAT/GRE (Graduate Record Examination) questions.[28] Allocations to the recipients were largest when recipients earned money and least when dictators earned money. Combining endowments earned by dictators with a double-blind payoff protocol has very striking results: in Cherry et al. (2002), 97 percent of the $40 dollar earners allocated $0 to the paired recipient, whereas the zero gift figure in Oxoby and Spraggon (2008) was 100 percent of dictators. In Oxoby and Spraggon's receiver earning treatment, 63 percent of dictators' offers exceeded 50 percent of the money for recipients who earned CAN$40 while no dictators made such offers in the (no earned endowments) baseline. Differences in generosity stemming from earned vs unearned endowments are also reported by others (e.g., Ruffle, 1998; List, 2007).

5 FOCAL POINTS: SCALE AND ASYMMETRY OF PAYOFFS

In a seminal paper, Schelling (1957) explores principles that underlie negotiations and bargaining (tacit or explicit) when communication is not possible. Focal points emerging in tacit (common- or divergent-interest) coordination games are thoroughly examined through a series of ingeniously constructed scenarios. Here we focus on common-interest and divergent-interest coordination games and the effect of scale of payoffs on the efficacy of focal points for reducing coordination failure.

The two games differ only with respect to the symmetry of payoffs when coordination is successful. With an action set of only two alternatives, say A and B, choosing differently is worth nothing. In the game with common interest, coordination on either action is valued the same $(\pi_i(A,A) = \pi_{\mathrm{i}}(B,B), i = a,b)$, whereas in the divergent-interest game players have favored actions: $\pi_a(A,A) > \pi_a(B,B)$ and $\pi_b(B,B) > \pi_b(A,A)$. The experimenter

27 This research program has taken on a life of its own: Google Scholar returns more than 60 000 results for "dictator games." A seminal paper in this literature is by Andreoni and Miller (2002) who vary the price of giving and dictators' endowment to test whether observed dictators' choices satisfy the generalized axiom of revealed preferences (GARP) (i.e., whether dictators' choices are rationalizable by convex preferences). They rule in favor of rationalizability, a conclusion that has been challenged by some later studies (List, 2007; Bardsley, 2008; Lazear, Malmendier and Weber, 2012; Cappelen, Nielsen and Sørensen, 2013).

28 In Cherry et al. (2002), subjects answering ten or more questions correctly received $40 endowments; others received $10 endowments. In Oxoby and Spraggon (2008), subjects who performed the earning task were assigned CAN$10, CAN$20 and CAN$40 if the total number of correct answers were fewer than nine, between nine and 14, or exceeding 15. See Engel (2011) for a meta-study of dictator games.

can induce these rankings by offering to pay the same positive payment in the common-interest game only when both players choose the same action. In the divergent-interest game, positive payments are again offered only when both players choose the same action but player *a*'s payment is larger if the coordination is on action *A* whereas player *b*'s payment is larger if the coordination is on action *B*. In the common-interest game, there are only two payoffs so we can interpret payoffs as von Neumann-Morgenstern (N-M) utilities. In contrast, in the divergent-interest game there are more than two distinct payoffs so we can only interpret induced payoffs as N-M utilities with the subsidiary hypothesis that the subjects are risk neutral.

Parravano and Poulsen (2015) study the effect of payoff scale on play in these two games.[29] If payoffs are indeed N-M utilities, then larger stakes through positive affine transformations of payoffs should have no effect on equilibrium play. But in the presence of decision costs, larger stakes may facilitate coordination by increasing the opportunity cost of not coordinating. In all cases, choosing different actions results in earnings of 0 for both players. However, the opportunity cost to each player of non-coordination is 30 (resp. ten) times as much in the High (resp. Medium) treatment than in the Low treatment. Authors find higher average coordination rates with High stakes than Low stakes but no difference between Medium stakes and Low stakes in the common-interest coordination game. Play in the divergent-interest game is not affected by the stakes. The authors argue that neither mixed strategies, nor Level-k reasoning, nor team reasoning can explain the pay effect on their data.

These data illustrate that if the opportunity cost of an inferior action is non-negligible and interdependent preferences are aligned with selfish ones (by design in the common-interest coordination game) then scaled-up payoff may be successful in promoting use of focal points[30] as coordination devices.

6 BINARY LOTTERY PAYOFFS FOR CONTROLLING CURVATURE OF UTILITIES

Consider an experiment in which subject j will be paid the amount of money $\pi(x_j, X_{-j})$ if they choose some action x_j and other agents choose the vector of actions X_{-j}. If the experimenter is testing the implications of a risk-neutral theory, they are interested in subject j addressing the experimental decision problem:

$$\max_{x_j} \int \pi(x_j, X_{-j}) f(X_{-j})\, dX_{-j} \tag{1.4}$$

where $f(X_{-j})$ represents risk-neutral agent j's belief that other risk-neutral agents will choose the vector of actions (X_{-j}). Let agent j be an expected utility (EU) maximizer and their preferences over money be represented by some increasing nonlinear function $u(\bullet)$. Then agent j has the subject's decision problem:

[29] In the symmetric coordination game, payoffs (in British pounds) are: (0.5, 0.5) in Low, (5, 5) in Medium and (15, 15) in High. In the asymmetric coordination game, payoffs are: (0.5, 0.6) in Low, (5, 6) in Medium and (15, 18) in High.

[30] See also Crawford, Gneezy and Rottenstreich (2008).

$$\max_{x_j} \int u(\pi(x_j, X_{-j}))\, g(X_{-j})\, dX_{-j} \tag{1.5}$$

where $g(X_{-j})$ represents EU agent j's belief that other EU agents will choose the vector of actions X_{-j}.

How could the experimenter incentivize agent j to take up the experimental decision problem? Binary lottery payoffs were introduced to the theoretical literature by Smith (1961) and to experimental methods by Roth and Malouf (1979). Let P denote the maximum possible payoff available to agent j in the game. Now, instead of paying the subject the positive amount of money $\pi(x_j, X_{-j})$ for action profile (x_j, X_{-j}) assign the agent prizes $A > B(\geq 0)$ with probabilities $\pi(x_j, X_{-j})/P$ and $1 - \pi(x_j, X_{-j})/P$. Without any loss of generality, set $u(A) = 1$ and $u(B) = 0$ and verify that by the axioms of expected utility (EU) theory the subject's decision problem would become:

$$\max_{x_j} \frac{1}{P} \int \pi(x_j, X_{-j})\, f(X_{-j})\, dX_{-j} \tag{1.6}$$

which is the same as the experimental decision problem.

Let us assume, for the discussion in this paragraph, that lottery payoffs are efficacious for controlling curvature of utility of payoffs in experiments. And consider for clarity of example that a researcher is testing a theoretical model that incorporates an assumption of linearity in payoffs (along with all the other assumptions in the model). Then, would an experimenter want to pay subjects with binary lottery payoffs rather than ordinary monetary payoffs? It depends on the experimenter's choice of research question. If a researcher wants to learn whether risk-neutralized subjects behave in ways consistent with the risk-neutral model's predictions then use of lottery payoffs would be warranted. But that might not be the researcher's question. The researcher may want to learn how well a model predicts the behavior of subjects with whatever risk attitudes and other uncontrolled characteristics they may have. Unless the intended domain of application of the model is limited to one in which all agents are known to be risk neutral, use of efficacious lottery payoffs might produce misleading conclusions about the usefulness of the model in predicting behavior.[31]

If we drop the assumption that lottery payoffs are known to be efficacious in inducing linearity in payoffs then questions about using them become more complicated. Papers reporting results from testing efficacy of lottery payoffs have produced conflicting conclusions.[32]

7 RESPONSE MODE EFFECTS CONFOUNDING INDUCED VALUATION

Suppose that an experimenter wants to elicit subjects' risk attitudes. The literature on the preference reversal phenomenon provides an example in which use of induced valuation

---

31 Whether or not a researcher "should" use lottery payoffs for experiments testing risk-neutral models has been disputed in an historical controversy: see Cox, Smith and Walker (1992) and Kagel and Roth (1992).

32 See for examples: Berg et al. (1986), Walker, Smith and Cox (1990), Rietz (1993), Cox and Oaxaca (1995), Berg et al. (2008), Stahl and Haruvy (2008), Harrison, Martínez-Correa and Swarthout (2013).

by the experimenter can systematically elicit different risk preferences from the subjects when the response mode is switched from choice to valuation. This literature also contains a clear example of how salient payoffs can interact with task repetition and economic institution to remove the response mode effect on elicited risk preferences.

The choice objects in preference reversal experiments are typically binary lotteries. Lotteries or "bets" are presented in pairs containing a probability bet and a money bet with similar expected values. The probability bet has a relatively high probability of a relatively low win state payoff. The money bet has a relatively low probability of a relatively high win state payoff. The experiments involve eliciting preferences with both the choice response mode and the valuation response mode. Valuations are elicited for minimum selling prices with an incentive-compatible revelation mechanism such as the Becker-DeGroot-Marschak (BDM: Becker, DeGroot and Marschak, 1964) mechanism. If the choice and valuation response modes elicit the same preferences, then, for any pair of lotteries, the money bet should have a higher selling price than the probability bet if and only if the money bet is chosen. The "preference reversal phenomenon" (PRP) is that the probability bet is frequently chosen when the money bet has a higher selling price.

Research on the PRP originated in psychology and was introduced into economics by Grether and Plott (1979). Their paper examines several reasons why the earlier results reported in psychology experiments might not have implications for economics. One of these reasons was use of hypothetical payoffs in the earlier literature.[33] Grether and Plott's treatments included use of hypothetical payoffs and (what were at the time) large monetary payoffs. They found (to their surprise) that incidence of preference reversals was somewhat higher in the monetary payoff treatments than in the otherwise identical hypothetical payoff experiments.

The central question addressed in the paper by Cox and Grether (1996) is whether the PRP is robust to repetitions in markets. They used the traditional BDM mechanism to elicit selling prices in non-market treatments. In market treatments, they used the second price sealed-bid auction and the English clock auction to elicit selling prices. In addition to experimenting with individual choice and market decision-making, Cox and Grether included treatments with hypothetical and real monetary payoffs of salient rewards. They also included treatments with and without repetition of tasks. Cox and Grether reported that:

- the PRP is robust to hypothetical vs monetary payoffs and to individual choice vs. market decisions *without* repetition;
- with hypothetical payoffs the PRP is robust to individual choice vs market decisions *with* repetition;
- with monetary payoffs, the PRP is *not* robust to five rounds of market decision-making.

A good question to pose for experimental economists is whether they have data that provides clear support for their shared methodological view that subjects should be paid salient monetary rewards. Results from this experiment provide an unambiguously clear

[33] One important exception to use of hypothetical payoffs was the experiment run on the floor of a Las Vegas casino by Lichtenstein and Slovic (1973).

example. The central question for the research is whether the PRP is robust to repeated decisions in markets. With data from hypothetical payoff treatments the answer is "yes." With data from money payoff treatments the answer is "no."

Insight into the reasons why hypothetical and financial incentives had these implications is provided by reviewing the features of the English clock auction in this experiment. Selling prices were elicited; therefore, the English clock displays decreasing prices. In the case of the money bet with \$16 win state payoff, the clock started at a price of \$16 and decreased by five cents every second. This means, for example, it would take 243 price clock ticks extending over 243 seconds for the price on the clock to decrease to this money bet's expected value of \$3.85. Subject impatience or disutility from time spent watching the price clock could be relieved by exiting the auction early in the price decrease time period. This is exactly what many subjects did in the treatment with hypothetical payoffs (Cox and Grether, 1996). But exiting early produces high selling prices for the dollar bet and results in high preference reversal rates. In contrast, in the treatments with financial incentives (\$16 in the win state or \$1.50 in the loss state, and 50 percent of these amounts in another treatment), subjects remained in the auction long enough to produce selling prices consistent with their choices.[34] This example supports an interpretation that it is important to avoid hypothetical payoffs when it is expected that subjects may get disutility from playing the game, and to use dominant salient monetary payoffs that are sufficiently high to compensate for the disutility. In contrast, when playing the game is valued, and accumulating hypothetical payoffs is enjoyable, as in some double auction experiments, paying real payoffs may not be necessary.

The PRP did not disappear in Cox and Grether's treatments that used repeated decision tasks and real payoffs when BDM was used to elicit selling prices. The efficacy of BDM in eliciting true buying or selling prices depends on subjects' understanding of the mechanism – an issue we take up next.

8 GAME-FORM RECOGNITION CONFOUNDING INDUCED VALUATION

A gap between willingness to pay (WTP) and willingness to accept (WTA) has been reported in 27 out of 39 experiments listed in Table 1 of Plott and Zeiler (2005).[35] The gap, however, disappears in Plott and Zeiler (2005), who use BDM as the elicitation mechanism and whose experimental procedure includes extensive training of subjects on the elicitation mechanism. In light of their findings, the authors advance the idea that "subject misconception" may explain the observed gap.[36]

Cason and Plott (2014) take the idea of "game misconception" one step further by studying the performance of the BDM mechanism in elicitation of a known value. In their experiment, the value of the object is induced (a card that can be exchanged for \$2), so no

[34] The win state payoff would be about \$26 in 2017 dollars. This was "real money" for undergraduate subjects.

[35] Endowment effect and loss aversion have appeared as favorite explanations of the WTP/WTA gap.

[36] For some controversy about these issues, see Isoni, Loomes and Sugden (2011) and Plott and Zeiler (2011).

"endowment" effect is expected (see also Kahneman, Knetsch and Thaler, 1990). Their data reveal a correlation between WTA and the upper bound of the support of stochastic prices used in the BDM elicitation.

Cason and Plott argue that low cost of "mistakes" from game misconception can explain this correlation but context-dependent "preferences" cannot. Subjects, in search of clues for understanding the new mechanism, may pool from their experience and (mistakenly) think of BDM as a first price (sellers') auction.[37] The "seller with the lowest ask sells the good and receives his or her own ask" is a common experience whereas exposure to exchanges happening at a price different than the seller's ask may be limited. If so, then the optimal asking price in the presence of such misconception is increasing in the upper bound of the support of BDM prices, which is consistent with the Cason and Plott data. These findings call for caution in interpreting data from experiments with possible game misconception as revealing non-standard preferences. Increasing the scale of payoffs is expected to fail in ameliorating this type of loss of control for subjects who are not aware of their own misconception and who experience little or no opportunity cost from their actions.

Almost half (17 out of 39) of the experiments from the literature on the gap of WTA/WTP reviewed by Plott and Zeiler (2005, Table 1) did not use an incentive-compatible mechanism in elicitation of valuations. The use of payoff protocols that are *not* incentive compatible is a common but puzzling practice by many researchers.

9 HOW NOT TO MISUSE INDUCED VALUATION: INCENTIVE-COMPATIBLE PAYOFF MECHANISMS

A payoff protocol that provides a mapping from a sequence of choices to realized payment is incentive *in*compatible if it provides incentives for a subject, who prefers option *a* over *b* in a single decision, to prefer *b* over *a* when the choice is embedded in a multiple decision setting. Use of incentive-*in*compatible payoff protocols is widespread in the literature, most especially in experiments on decisions under risk but also in many other topic areas (Cox et al., 2015).

The experiment on the preference reversal phenomenon by Grether and Plott (1979) generated methodological critiques that have not been generally appreciated more than three decades later. Grether and Plott (1979, p. 623) stated that "this behavior is not simply a violation of some type of expected utility hypothesis. . . It suggests that no optimization principles of any sort lie behind even the simplest of human choices." Such claims of generality elicited critiques by Holt (1986) and Karni and Safra (1987). Karni and Safra (1987) explained that the BDM mechanism used by Grether and Plott to elicit selling prices for the binary lottery options in the experiment requires reduction and independence axioms to interpret data as preference reversals.

Grether and Plott asked their subjects to make several decisions and randomly selected one decision for payoff. They did this to control for portfolio effects from realizing payoffs from more than one decision at the end of the experiment and for wealth effects from

[37] Authors identify 111 (out of 244) subjects whose choices suggest this type of misconception.

paying each decision immediately after it was made.[38] Their random selection payoff protocol has been given several names in the literature; perhaps the most commonly used name is random lottery incentive mechanism (RLIM). Holt (1986) explained that RLIM depends on the reduction and independence axioms from expected utility theory for interpretation of the data as preference reversals. The reasoning is as follows. The choice options in the experiment are binary lotteries. Randomly selecting one decision for payoff is itself a lottery. Hence, the incentives provided to the subjects are compound lotteries consisting of the payoff mechanism "wrapped around" the binary lottery options in the experiment. Interpretation of the data as revealing preferences over the binary lotteries themselves, rather than preferences over the compound lottery used to incentivize the subjects, depends on the reduction and independence axioms; without these axioms, there is no theoretical basis for identifying a subject's responses as revealing their true preferences over the binary lotteries in the experiment.[39]

The critique by Holt (1986) created a problem for experimental methods intended to test theories of decision-making under risk other than expected utility theory. Absent known incentive-compatible payoff protocols, much literature over the subsequent three decades simply ignores Holt's critique and applies RLIM in experiments intended to test alternatives to expected utility theory such as rank-dependent utility theory and cumulative prospect theory. Serious problems with this approach are provided by counterexamples to incentive compatibility of RLIM in the absence of the independence axiom (Holt, 1986; Cox et al., 2015) and empirical data demonstrating that the bias introduced by this payoff mechanism is significant (Cox, Sadiraj and Schmidt, 2014, 2015; Harrison and Swarthout, 2014).

As first asserted by Kahneman and Tversky (1979), a common justification for using RLIM in the absence of the independence axiom is the empirical assertion that subjects isolate each choice independently of the other choice opportunities in an experiment. This assertion is appealing to experimentalists because it simplifies experimental design: it allows one to ignore complications in interpreting data that come from possible cross-task contamination in many-decision experiments. There have been some papers that report isolation is supported by experimental data. Starmer and Sugden (1991) and Hey and Lee (2005a) consider the two extreme hypotheses that subjects: (1) isolate each choice; or

[38] To illustrate "portfolio" and "wealth" effects consider two pairs of lotteries $\{a, b\}$ and $\{c, d\}$. Let the experimental decision problem be ranking of lotteries in each pair. The experimenter asks the subject to choose between a and b and also choose between d and c. *Portfolio effect*: suppose that the experimenter pays the independently realized outcome from each chosen lottery at the end of the experiment. In that case, the subject's decision problem is to choose the most preferred out of the four available portfolios of two lotteries: $\{[a, c], [a, d], [b, c]$ or $[b, d]\}$. The experimental and subject's decision problems are not the same. For example, if $\{a, b\}$ and $\{c, d\}$ are the same pair then an EU individual with power 0.5 utility function prefers option a (30 for sure) over option b (100 or 0 with even odds) but he prefers portfolio $[b, b]$ if he is asked to choose from $\{a, b\}$ twice. *Wealth effect*: now suppose that the experimenter realizes the payoff from the first chosen lottery in pair $\{a, b\}$ right after the choice is made. If the outcome is some x then when the subject chooses again her choice set is $\{x + c, x + d\}$ and therefore, the experimental and subject's decision problems are not the same. This is known in the literature as "wealth effect." Using the example utility above, the subject chooses b the first time. If the outcome is 100 then the second time he chooses b again but if the outcome is 0 then he chooses a.

[39] Cox et al. (2015) offer counterexamples to incentive compatibility of RLIM in the absence of the independence axiom. Some wording in Azrieli, Chambers and Healy (2018) might suggest to readers that RLIM (or random problem selection – RPS – in their terms) is incentive compatible for decision theories other than expected utility theory. But their "Fact 1.1" (Azrieli et al., 2018, p. 1488) makes clear their agreement with earlier work that the independence axiom is required, given reduction of compound lotteries.

(2) make all choices so as to yield the most preferred probability distribution of payoffs from the whole experiment (which they named "full reduction"). They conclude that full reduction can be rejected in favor of isolation. Hey and Lee (2005b) test isolation against a hypothesis of "partial reduction" in which the current decision task can be given higher or lower weight than preceding tasks when a subject makes decisions so as to choose the best probability distribution for the whole experiment. The problem with this approach is that there are many alternatives to isolation other than full or partial reduction, including the more plausible hypothesis of cross-task contamination between choices.[40]

Recent empirical tests have produced consistent evidence of cross-task contamination from RLIM. Cox et al. (2014) report that asymmetrically dominated options can be introduced into sets of choice options in ways that permit systematic manipulation of subjects' choices when they are paid with RLIM.[41] Cox et al. (2015) report additional tests for cross-task contamination by three versions of RLIM and find that their RLIM data are characterized by significant choice order effects on revelation of classical paradoxical patterns. Harrison and Swarthout (2014) report an experiment in which a treatment with a single decision is used to generate control data for use in econometric analysis of risk attitudes with data generated by a treatment with many decisions and use of RLIM to select one decision for payoff. They find that estimated probability weighting for rank-dependent utility is biased by data generated with RLIM.

Methodological problems with use of incentive-incompatible payoff protocols are not confined to experiments on theories of decision under risk; instead, as explained by Cox et al. (2015), these problems are present in many other topic areas including research on social preferences, bargaining, public goods, and voting. The continuing problem with use of incentive-incompatible payoff protocols apparently stems, in part, from paucity of payoff protocols (for use in multiple decision experiments) that do not require the independence axiom. Partial solutions to this problem are beginning to be discovered.

Cox et al. (2015) report theoretical and empirical properties of a new payoff protocol – pay all correlated (PAC) – that is, incentive compatible for the dual theory of expected utility (Yaari, 1987).[42] Li (2016) introduced the accumulative best choice (ABC) mechanism that is incentive compatible for general risk theories, regardless of whether they include the independence axiom, which can be used in some types of applications.[43] Li finds that ABC is also behaviorally incentive compatible for her data.

---

[40] Starmer and Sugden (1991, p. 977) report one test for cross-task contamination in which the two-sided p-value is 0.051.

[41] Let the choice set, S_1 be a pair of options $\{A, B\}$. Construct a new choice set, $S_2 = \{C, D\}$ such that A dominates C but D dominates B. Choosing from S_2 first increases attractiveness of A in a subsequent choice from S_1. In one treatment in Cox et al. (2014), A and C were safe options with payments €4 and €3, whereas B and D were even-odds risky lotteries with prizes 0 or, resp., €10 and €12. Option B was chosen by 83 percent of subjects in a single decision task (S_1), whereas in the two decision tasks (first S_2, then S_1) only 52 percent of subjects chose B.

[42] With the PAC mechanism, outcomes from all chosen lotteries are paid for one realization of the state of nature at the end of the experiment. For this mechanism, states of nature need to be defined and all lotteries arranged to be comonotonic. That is, let states be indexed by $s(=1, \ldots, m)$ and lotteries by $i(=1, \ldots, n)$. If we let π_{is} denote the payoff from lottery i in state s then $\pi_{is} \geq \pi i(s+1)$ for all states $s(< m)$ and all lotteries i.

[43] With the accumulated best choice mechanism, the chosen option in decision round k is one of the feasible options in round $k+1$. The subject is paid according to the option chosen in the final round.

9.1 Summary Implications

Assuming that the researcher wants to use payoff protocols that are incentive compatible − in other words that they want to avoid using payoff protocols that create incentives for biased choices − then they should observe the following principles of experimental methods:[44]

- For risk-neutral models, all commonly used methods of paying subjects are incentive compatible.
- Independently realizing outcomes for more than one decision at the end of an experiment creates portfolio incentives for risk-averse agents.
- Randomly selecting one decision for payoff is not incentive compatible unless the independence axiom is assumed.
- Paying all decisions sequentially is incentive compatible for decision theories defined over income, not accumulated wealth.
- For comonotonic options, paying all choices at the end of the experiment with one realization of the state of nature is incentive compatible for the dual theory of expected utility.
- Giving each subject only one decision task is incentive compatible for all decision theories.
- The accumulated best choice (ABC) mechanism is incentive compatible for decision theories with complete and transitive preferences.

10 CONCLUDING REMARKS

Incentives are a major part of what defines the "economics" part in experimental economics. We have discussed several experimental approaches to addressing confounds identified in the literature. They include: choice of salient payoff levels to insure dominance, experience (practice rounds, repetition) to promote understanding of decision tasks and unfamiliar institutions, withholding information on others' payoffs (privacy) to deter interpersonal preferences (when appropriate), alternative institutional formats (e.g., game trees and price clocks) to vary accessibility of payoff information; and lottery payoffs to induce risk neutrality. We have examined issues that arise with response mode effects and game-form misconception. Choice of which instruments to use and when to use them is part of the art of experimental design that should be informed by the research questions of interest and the type of confounds that are expected in a given environment and economic institution. One confound that is frequently not addressed in the literature is use of payoff protocols that are not incentive compatible. We call attention to this issue.

[44] We end this section with these summary statements because widely read journals continue to publish papers reporting experiments with unknown bias in data from use of payoff mechanisms known to be incentive incompatible. Cox et al. (2015) contains examples from papers in multiple journals on a range of topic areas.

REFERENCES

Andersen, Steffen, Seda Ertaç and Uri Gneezy et al. 2011. "Stakes matter in ultimatum games." *American Economic Review* **101**(7): 3427–39.

Andreoni, James and John Miller. 2002. "Giving according to GARP: an experimental test of the consistency of preferences for altruism." *Econometrica* **70**(2): 737–53.

Ariely, Dan, Uri Gneezy, George Loewenstein and Nina Mazar. 2009. "Large stakes and big mistake." *Review of Economic Studies* **76**(2): 451–69.

Aumann, Robert J. 1992. "Irrationality in game theory." In Partha Dasgupta, Douglas Gale, Oliver Hart and Eric Maskin (eds), *Economic Analysis of Markets and Games* (pp. 214–27). Cambridge, MA: MIT Press.

Azrieli, Yaron, Christopher P. Chambers and Paul J. Healy. 2018. "Incentives in experiments: a theoretical analysis." *Journal of Political Economy* **126**(4): 1472–503.

Bardsley, Nicholas. 2008. "Dictator game giving: altruism or artefact?" *Experimental Economics* **11**(2): 122–33.

Becker, Gordon M., Morris H. DeGroot and Jacob Marschak. 1964. "Measuring utility by a single-response sequential method." *Behavioral Science* **9**(3): 226–32.

Berg, Joyce E., Thomas A. Rietz and John W. Dickhaut. 2008. "On the performance of the lottery procedure for controlling risk preferences." In Charles R. Plott and Vernon L. Smith (eds), *Handbook of Experimental Economics, Results Vol. 1* (pp. 1087–97). Amsterdam: North-Holland.

Berg, Joyce E., Lane A. Daley, John W. Dickhaut and John R. O'Brien. 1986. "Controlling preferences for lotteries on units of experimental exchange." *Quarterly Journal of Economics* **101**(2): 281–306.

Bolton, Gary E. 1991. "A comparative model of bargaining: theory and evidence." *American Economic Review* **81**(5): 1096–136.

Camerer, Colin F. and Robin M. Hogarth. 1999. "The effects of financial incentives in experiments: a review and capital–labor–production framework." *Journal of Risk and Uncertainty* **19**(1): 7–42.

Cameron, Lisa A. 1999. "Raising the stakes in the ultimatum game: experimental evidence from Indonesia." *Economic Inquiry* **37**(1): 47–59.

Cappelen, Alexander W., Ulrik H. Nielsen and Erik Ø. Sørensen et al. 2013. "Give and take in dictator games." *Economics Letters* **118**(2): 280–83.

Cason, Timothy N. and Charles R. Plott. 2014. "Misconceptions and game form recognition: challenges to theories of revealed preference and framing." *Journal of Political Economy* **122**(6): 1235–70.

Cherry, Todd L., Peter Frykblom and Jason F. Shogren. 2002. "Hardnose the dictator." *American Economic Review* **92**(4): 1218–21.

Cox, James C. and David M. Grether. 1996. "The preference reversal phenomenon: response mode, markets and incentives." *Economic Theory* **7**(3): 381–405.

Cox, James C. and Duncan James. 2012. "Clocks and trees: isomorphic Dutch auctions and centipede games." *Econometrica* **80(**2): 883–903.

Cox, James C. and Ronald L. Oaxaca. 1995. "Inducing risk-neutral preferences: further analysis of the data." *Journal of Risk and Uncertainty* **11(**1): 65–79.

Cox, James C., Vjollca Sadiraj and Ulrich Schmidt. 2014. "Asymmetrically dominated choice problems, the isolation hypothesis and random incentive mechanisms." *PloS ONE* **9**(3): e90742.

Cox, James C., Vjollca Sadiraj and Ulrich Schmidt. 2015. "Paradoxes and mechanisms for choice under risk." *Experimental Economics* **18**(2): 215–50.

Cox, James C., Vernon L. Smith and James M. Walker. 1992. "Theory and misbehavior of first-price auctions: comment." *American Economic Review* **82**(5): 1392–412.

Crawford, Vincent P., Uri Gneezy and Yuval Rottenstreich. 2008. "The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures." *American Economic Review* **98**(4): 1443–58.

Croson, Rachel T.A. 2000. "Thinking like a game theorist: factors affecting the frequency of equilibrium play." *Journal of Economic Behavior & Organization* **41**(3): 299–314.

Engel, Christoph. 2011. "Dictator games: a meta study." *Experimental Economics* **14**(4): 583–610.

Falk, Armin, Stephan Meier and Christian Zehnder. 2014. "Do lab experiments misrepresent social preferences? The case of self-selected student samples." *Journal of the European Economics Association* **11**(4): 839–52.

Fey, Mark, Richard D. McKelvey and Thomas R. Palfrey. 1996. "An experimental study of constant-sum centipede games." *International Journal of Game Theory* **25**(3): 269–87.

Fouraker, Lawrence E. and Sidney Siegel. 1963. *Bargaining Behavior*. New York: McGraw-Hill.

Grether, David M. and Charles R. Plott. 1979. "Economic theory of choice and the preference reversal phenomenon." *American Economic Review* **69**(4): 623–38.

Güth, Werner, Rolf Schmittberger and Bernd Schwarze. 1982. "An experimental analysis of ultimatum bargaining." *Journal of Economic Behavior & Organization* **3**(4): 367–88.

Harrison, Glenn W. and J. Todd Swarthout. 2014. "Experimental payment protocols and the bipolar behaviorist." *Theory and Decision* **77**(3): 423–38.

Harrison, Glenn W., Jimmy Martínez-Correa and J. Todd Swarthout. 2013. "Inducing risk neutral preferences with binary lotteries: a reconsideration." *Journal of Economic Behavior & Organization*, **94**: 145–59.
Hertwig, Ralph and Andreas Ortmann. 2001. "Experimental practices in economics: a methodological challenge for psychologists?" *Behavioral and Brain Sciences* **24**(3): 383–451.
Hertwig, Ralph and Andreas Ortmann. 2003. "Economists' and psychologists' experimental practices: how they differ, why they differ, and how they could converge." In Isabelle Brocas and Juan D. Carrillo (eds), *The Psychology of Economic Decisions* (pp. 253–72). Oxford: Oxford University Press.
Hey, John D. and Jinkwon Lee. 2005a. "Do subjects separate (or are they sophisticated)?" *Experimental Economics* **8**(3): 233–65.
Hey, John D. and Jinkwon Lee. 2005b. "Do subjects remember the past?" *Applied Economics* **37**(1): 9–18.
Hoffman, Elizabeth, Kevin McCabe, Keith Shachat and Vernon L. Smith. 1994. "Preferences, property rights, and anonymity in bargaining games." *Games and Economic Behavior* **7**(3): 346–80.
Holt, Charles A. 1986. "Preference reversals and the independence axiom." *American Economic Review* **76**(3): 508–15.
Isoni, Andrea, Graham Loomes and Robert Sugden. 2011. "The willingness to pay–willingness to accept gap, the 'endowment effect,' subject misconceptions, and experimental procedures for eliciting valuations: comment." *American Economic* Review **101**(2): 991–1011.
Kagel, John H. and Alvin E. Roth. 1992. "Theory and misbehavior in first-price auctions: comment." *American Economic Review* **82**(5): 1379–91.
Kahneman, Daniel and Amos Tversky. 1979. "Prospect theory: an analysis of decision under risk." *Econometrica* **47**(2): 263–91.
Kahneman, Daniel, Jack L. Knetsch and Richard H. Thaler. 1986. "Fairness and the assumptions of economics." *Journal of Business* **59**(4): S285–S300.
Kahneman, Daniel, Jack L. Knetsch and Richard H. Thaler. 1990. "Experimental tests of the endowment effect and the Coase theorem." *Journal of Political Economy* **98**(6): 1325–48.
Karagözoglu, Emin and Ümit Barıs Urhan. 2017. "The effect of stake size in experimental bargaining and distribution games: a survey." *Group Decision and Negotiation* **26**(2): 285–325.
Karni, Edi and Zvi Safra. 1987. "'Preference reversal' and the observability of preferences by experimental methods." *Econometrica* **55**(3): 675–85.
Krockow, Eva M., Andrew M. Colman and Briony D. Pulford. 2016. "Exploring cooperation and competition in the centipede game through verbal protocol analysis." *European Journal of Social Psychology* **46**(6): 746–61.
Lazear, Edward P., Ulrike Malmendier and Roberto A. Weber. 2012. "Sorting in experiments with application to social preferences." *American Economic Journal: Applied Economics* **4**(1): 136–63.
Levitt, Steven D., John A. List and Sally E. Sadoff. 2011. "Checkmate: exploring backward induction among chess players." *American Economic Review* **101**(2): 975–90.
Li, Yi. 2016. "Incentive compatible payoff mechanisms for general risk theories." Accessed 4 February, 2019 at http://scholarworks.gsu.edu/econ_diss/120.
Lichtenstein, Sarah and Paul Slovic. 1973. "Response-induced reversals of preference in gambling: an extended replication in Las Vegas." *Journal of Experimental Psychology* **101**(1): 16–20.
List, John A. 2007. "On the interpretation of giving in dictator games." *Journal of Political Economy* **115**(3): 482–93.
McKelvey, Richard D. and Thomas R. Palfrey. 1992. "An experimental study of the centipede game." *Econometrica* **60**(4): 803–36.
Munier, Bertrand and Costin Zaharia. 2002. "High stakes and acceptance behavior in ultimatum bargaining." *Theory and Decision* **53**(3): 187–207.
Oosterbeek, Hessel, Randolph Sloof and Gijs van de Kuilen. 2004. "Cultural differences in ultimatum game experiments: evidence from a meta-analysis." *Experimental Economics* **7**(2): 171–88.
Oxoby, Robert J. and John Spraggon. 2008. "Mine and yours: property rights in dictator games." *Journal of Economic Behavior & Organization* **65**(3–4): 703–13.
Palacios-Huerta, Ignacio and Oscar Volij. 2009. "Field centipedes." *American Economic Review* **99**(4): 1619–35.
Parravano, Melanie and Odile Poulsen. 2015. "Stake size and the power of focal points in coordination games: experimental evidence." *Games and Economic Behavior* **94**(4): 191–9.
Plott, Charles R. and Kathryn Zeiler. 2005. "The willingness to pay–willingness to accept gap: the 'endowment effect,' subject misconceptions, and experimental procedures for eliciting valuations." *American Economic Review* **95**(3): 530–45.
Plott, Charles R. and Kathryn Zeiler. 2011. "The willingness to pay–willingness to accept gap, the 'endowment effect,' subject misconceptions, and experimental procedures for eliciting valuations: reply." *American Economic Review* **101**(2), 1012–28.
Rabin, Matthew. 1993. "Incorporating fairness into game theory and economics." *American Economic Review* **83**(5): 1281–302.
Rapoport, Amnon, William E. Stein, James E. Parco and Thomas E. Nicholas. 2003. "Equilibrium play and adaptive learning in a three-person centipede game." *Games and Economic Behavior* **43**(2), 239–65.

Rietz, Thomas A. 1993. "Implementing and testing risk-preference-induction mechanisms in experimental sealed-bid auctions." *Journal of Risk and Uncertainty* **7**(2): 199–213.
Rosenthal, Robert W. 1981. "Games of perfect information, predatory pricing and the chain-store paradox." *Journal of Economic Theory* **25**(1): 92–100.
Roth, Alvin E. and Michael W.K. Malouf. 1979. "Game-theoretic models and the role of information in bargaining." *Psychological Review* **86**(6): 574–94.
Ruffle, Bradley J. 1998. "More is better, but fair is fair: tipping in dictator and ultimatum games." *Games and Economic Behavior* **23**(2): 247–65.
Schlag, Karl H., James Tremewan and Joël J. van der Weele. 2015. "A penny for your thoughts: a survey of methods for eliciting beliefs." *Experimental Economics* **18**(3): 457–90.
Schelling, Thomas C. 1957. "Bargaining, communication, and limited war." *Conflict Resolution* **1**(1): 19–36.
Selten, Reinhard. 1965. "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit" [Game-theoretical treatment of an oligopoly model with demand inertia]. *Zeitschrift für die gesamte Staatswissenschaft* **121**(2): 301–24 and 667–89.
Siegel, Sidney. 1961. "Decision making and learning under varying conditions of reinforcement." *New York Academy of Sciences Annals* **89**(5): 766–83.
Slonim, Robert and Alvin E. Roth. 1998. "Learning in high stakes ultimatum games: an experiment in the Slovak Republic." *Econometrica* **66**(3): 569–96.
Smith, Cedric A.B. 1961. "Consistency in statistical inference and decision." *Journal of the Royal Statistical Society. Series B (Methodological)* **23**(1): 1–37.
Smith, Vernon L. 1976. "Experimental economics: induced value theory." *American Economic Review* **66**(2): 274–9.
Smith, Vernon L. 1982a. "Microeconomic systems as an experimental science." *American Economic Review* **72**(5): 923–55.
Smith, Vernon L. 1982b. "Markets as economizers of information: experimental examination of the 'Hayek hypothesis'." *Economic Inquiry* **20**(2): 165–79.
Smith, Vernon L. 2010. "Theory and experiment: what are the questions?" *Journal of Economic Behavior & Organization* **73**(1): 3–15.
Smith, Vernon L. and James M. Walker. 1993. "Monetary rewards and decision cost in experimental economics." *Economic Inquiry* **31**(2), 245–61.
Stahl, Dale O. and Ernan Haruvy. 2008. "Subgame perfection in ultimatum bargaining trees." *Games and Economic Behavior* **63**(1): 292–307.
Starmer, Chris and Robert Sugden. 1991. "Does the random-lottery incentive system elicit true preferences? An experimental investigation." *American Economic Review* **81**(4), 971–8.
Tisserand, Jean-Christian. 2016. "Le jeu de l'ultimatum, une méta-analyse de 30 années de recherches expérimentales" [The ultimatum game, a meta-analysis of 30 years of experimental research]. *L'Actualité économique* **92**(1–2): 289–314.
Walker, James M., Vernon L. Smith and James C. Cox. 1990. "Inducing risk-neutral preferences: an examination in a controlled market environment." *Journal of Risk and Uncertainty* **3**(1): 5–24.
Wilde, Louis L. 1981. "On the use of laboratory experiments in economics." In Joseph C. Pitt (ed.), *Philosophy in Economics* (pp. 137–48). Dordrecht: Springer.
Yaari, Menahem E. 1987. "The dual theory of choice under risk." *Econometrica* **55**(1): 95–115.