Journal of Economic Behavior & Organization 131 (2016) 141–150

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

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

Experimental methods: Pay one or pay all

Gary Charness[a,*], Uri Gneezy[b], Brianna Halladay[a]

[a] Department of Economics, University of California, Santa Barbara, United States
[b] Rady School of Management, University of California, San Diego, United States, and CREED, University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 15 March 2016
Received in revised form 23 July 2016
Accepted 26 August 2016
Available online 31 August 2016

JEL Classifications:
B49
C91
C99

Keywords:
Experiments
Payment approaches
Incentives

ABSTRACT

In some experiments participants make multiple decisions; this feature facilitates gathering a considerable amount of incentivized data over the course of a compact session. A conservative payment scheme is to pay for the outcome from every decision made. An alternative approach is to pay for the outcome of only a subset of the choices made, with the amount at stake for this choice multiplied to compensate for the decreased likelihood of that choice's outcome being drawn for payoff. This "pay one" approach can help to avoid wealth effects, hedging, and bankruptcy considerations. A third method is to pay only a subset of the participants for their choices, thereby minimizing transactions costs. While the evidence on differences across payment methods is mixed, overall it suggests that paying for only a subset of periods or individuals is at least as effective as the "pay all" approach and can well be more effective. We further the discussion about how to best choose an incentive structure when designing an experiment.

1. Introduction

What distinguishes laboratory experiments in economics from most experiments in other social sciences is the use of financial incentives. The working assumption is that providing incentives for choices and outcomes leads to more meaningful and reliable choice behavior. Specifically, incentives that encourage subjects to make honest and non-arbitrary choices will accurately reveal characteristics of their pReferences

The simplest and cleanest approach is to have only a single experimental choice. Cross-task contamination is not an issue with single-choice experiments and there is no way for subjects to hedge decisions. However, there are also a number of limitations to this approach. While the incentive structure in single-choice experiments is extremely salient to the subjects, this type of experiment only allows for between-subject comparisons, requiring a significant amount of observations and relatively high cost. In addition, it is impossible to study learning or to examine more settled behavior in relation to equilibrium predictions in an experiment involving some complexity. Thus, it is common in experiments for participants to make multiple decisions or one decision for each of a number of periods. This approach is useful for gathering a considerable amount of incentivized data over the course of a compact session, allowing the researcher to understand individuals' fixed effects and to observe learning.

The traditional and conservative approach in experiments with multiple decisions is to pay for the outcome from every decision made ("pay all"). An alternative approach is to pay for only one of the choices made, drawn at random at the end of

* Corresponding author.
E-mail addresses: charness@econ.ucsb.edu (G. Charness), ugneezy@ucsd.edu (U. Gneezy), briannahalladay8@gmail.com (B. Halladay).

http://dx.doi.org/10.1016/j.jebo.2016.08.010
0167-2681/

the session ("pay one"). When using pay one, the amount at stake per choice is multiplied to compensate for the decreased likelihood of that choice's outcome being drawn for payoff. A related approach to pay one is to pay for a selected number of periods rather than just one. Doing so helps to even out the earnings made by participants, which can be useful in terms of subject-pool maintenance. For example, one might pay for one period of every 10, so that five periods are chosen for payment in a 50-period session.

In terms of current practice, none of the 15 experiments with multiple decisions reported in the top-five economics journals in 2011 used pay one, and it appears that, overall, pay one is used less than half as frequently as pay all (Azrieli et al., 2015). Given this current practice, we consider pay all to be the baseline approach in this paper. Yet, while it appears that the majority of well-published multi-period experiments are implementing pay all, some recent research suggests a theoretical advantage of pay one (Azrieli et al., 2015). In this paper we discuss the two approaches, present the challenges and benefits of each, and discuss how the relative advantage changes depending on the particular laboratory setting. Experimenters must weigh a variety of considerations when choosing the incentive structure of their projects.

Paying for all decisions makes wealth and portfolio effects possible, as well as cross-task contamination. In experiments where losses are possible, this approach also runs the risk of having bankruptcy issues distort incentives: If a participant knows that she has a negative aggregate payoff during a session and knows that losses cannot be enforced, it is in her interest to take risks in the hope of attaining positive payoffs. In multi-period experiments where bankruptcy issues are feasible, paying for a period or periods determined after the experiment ameliorates or eliminates this issue.[1]

Consider instead a multi-decision experiment where subjects earn money for a single randomly-selected decision.[2] This incentive structure eliminates the opportunity for wealth and portfolio effects and eliminates hedging opportunities (Bardsley et al., 2009). If one holds the expected payment for a decision the same across pay one and pay all, this means that the nominal payment will be correspondingly larger (e.g., 10 times as large if one of 10 decisions will be paid).[3]

Another form of this question is whether paying only a subset of the people (rather than choices) in an experiment induces different choice behavior than paying every participant. Paying only a subset can help with logistical considerations, for example with online experiments and experiments conducted in large classes, since far fewer monetary transactions need be made in this case. Can this effort-saving approach be used without distorting choices compared to those made with a pay-every-person approach?

In this article we explore the issues involving each of these incentive structures. We present evidence for and against each incentive option in diverse laboratory settings, and mention some recent and related ideas. We aim at providing a reference and guide for experimental economists. We present evidence regarding pay one and pay all in Section 2 and present evidence regarding paying all subjects or a subset in Section 3. We conclude in Section 4.

2. Pay one or pay all: evidence

The theoretical literature regarding pay one or pay all in experiments delivers predictions that depend on assumptions about a specific utility theory and the environment. See for example Holt (1986) and Karni and Safra (1987) regarding risk and expected utility, Chandrasekhar and Xandri (2014) regarding stochastic termination, and Baillon et al. (2015) who show theoretically that the order of the randomization and the event is important in ambiguity experiments. In a recent paper, Azrieli et al. (2015) conclude that in a rather general environments that include strategic choices paying for only one period (and in fact, not having a show-up fee paid) is the best mechanism.

The general conclusion from this theoretical literature is that the benchmark prediction is sensitive to the assumptions made. Hence, for theoretically-motivated experiments, the incentive structure should be part of the benchmark prediction of the model, and the equivalence between the two payments should not be assumed. However, the focus of this paper is not on the theoretical aspect of the comparison between pay all and pay one, but rather on practical considerations in the laboratory. We now proceed to empirical evidence regarding each incentive structure, specifying the laboratory setting.

Table 1 summarizes the results of each study discussed.

[1] For example, Charness and Levin (2009) examine the winner's curse (in the "Acquire-a-Company" task) in 60-period sessions. Bidding positive amounts can generate losses even larger than the endowment given. To mitigate the bankruptcy issue, smooth the overall individual payoffs to some extent, and maintain the subjects' attention throughout a session, one period from each 10-period range was chosen for actual payoff at the end of the session.

[2] Cubitt et al. (2001) note that if randomization is involved, it is imperative that subjects understand the process by which this is done. A related potential concern with pay one is the introduction of "background risk," discussed in section 3.

[3] It is an open empirical question if and when having a smaller-but-certain nominal payoff leads to more attentiveness than a larger-but-less likely nominal payoff. Anecdotal evidence from Cabrales et al. (2003) shows that in a task requiring considerable cognitive resources, when the nominal payoff was 50, 100, or 150 and each period was paid, some people made arbitrary responses (i.e., choosing the same "color" in all 10 periods, despite different circumstances). However, when the nominal payoff was instead 500, 1000, or 1500 (with only a 10% chance of that period being selected for payment), no such behavior was observed. This suggests that the higher nominal payoff is more salient than the reduction in the likelihood of payment for a particular period.

Table 1
Results of experimental studies.

Author(s)	Experimental Task	Summary of Results
Andersen et al. (2014)	Discounting task	The effect of paying only a subset of subjects by exogenously varying the probability of payment from 10% to 100%. Do not find an effect on subject behavior
Baltussen et al. (2012)	Deal or No Deal (dynamic setting)	Risk preferences were consistent on average across the pay one and single-choice treatments. However, in the treatment where only 10% of subjects were selected at random for payment, observe a significant decrease in risk aversion.
Bardsley et al. (2009)	Book Chapter about incentive mechanisms	Indicate that the pay one incentive structure eliminated the opportunity for wealth and portfolio effects and eliminates hedging opportunities. They find little evidence of bias when using the random lottery incentive scheme, and conclude there is more support than opposition for this payment scheme.
Beaud and Willinger (2015)	Risk elicitation	Observe no difference in risk vulnerability whether only 10% of subjects are selected at random for payment or all subjects are paid.
Blanco et al. (2010)	Cooperation game with belief elicitation	Do not find evidence of hedging when paying for both the game outcome and the belief elicitation. Additionally, stated beliefs did not differ whether subjects were paid for both the game outcome and the belief elicitation or one selected at random.
Bolle (1990)	Social Preferences	No difference in behavior across the treatment where subjects are paid for all decisions (pay all) and the treatment where two out of twenty subjects are randomly selected for payment.
Brokesova et al. (2015)	Risk elicitation	Observe no difference in behavior on a risk-taking task between a single choice, a task selected at random for payment, and a subject selected at random for payment.
Casari et al. (2007)	First price sealed-bid common value auctions	Inexperienced bidders with larger cash balances bid somewhat less aggressively than those with smaller balances. Argue that the effect on behavior is not economically significant but they find even less influence on behavior with experienced bidders.
Clot et al. (2015)	Social Preferences	Find no difference in behavior when all participants are paid versus when only some participants are paid.
Cox (2009)	Social Preferences	Stronger social context (repeated interaction with pay one) resulted in more generous behavior by dictators, second movers in the investment game but not first movers in the investment game than with weak social context (single-choice)
Cox et al. (2008)	Social Preferences	Conduct a test of trust, fear, and reciprocity, with both a single-choice treatment and a pay one treatment and do not observe differences in subject behavior across treatments.
Cox et al. (2014)	Lottery Choice	Subjects choose the safe option more frequently when listed next to an asymmetrically dominated alternative than when presented in isolation, resulting in a preference reversal that creates concern about using pay one.
Cox and Sadiraj (2015)	Lottery Choice	Test eight different incentive mechanisms and observe large differences in revealed risk preferences across the eight options.
Cubitt et al. (1998)	Lottery Choice	Find no evidence of a contamination effect and cannot reject the independence axiom therefore behavior appears unbiased with pay one. Additionally, there is no evidence that behavior under the pay one incentive structure is less risk averse.
Erat and Gneezy (2012)	Deception game	Some people prefer not to lie even when lying increases theirs and the receiver's payoff. One out of twenty subjects were paid.
Garvin and Kagel (1994)	First price sealed-bid common value auctions	Find that the coefficient on the cash balance to be statistically significant and negative, indicating evidence that paying all can have a significant effect on subject bidding behavior.
Ham et al. (2005)	First price sealed-bid private value auctions	Accumulated cash balances significantly affect behavior and subjects seem to have income targets. Subjects bid high initially to try and win the auction but as cash balances rise and subjects approach what appears to be an income target, bids fall. This suggests that here pay all may have a problematic effect on behavior.
Harrison et al. (2007)	MPL	Do not observe a difference in subject behavior when paying only 10% of the subjects versus paying all subjects.
Harrison et al. (2013)	Lottery Choice	Facing multiple lottery choices and pay one, subjects' choices show a significant shift towards risk-neutral behavior, in line with Expected Utility Theory .
Harrison and Swarthout (2014)	Lottery Choice	When implementing pay one, compared to a single-choice experimental design, there is no difference in behavior under the assumption of Expected Utility Theory.
Laury (2006)	MPL	Finds risk aversion to be highest under high stakes but does not observe a difference in behavior when comparing pay all and pay one under low stakes, therefore, concludes that pay one is not causing subjects to scale down incentives to account for the random selection.

Table 1 (*Continued*)

Author(s)	Experimental Task	Summary of Results
Lee (2008)	Change improving decision model	The introduction of background risk due to the pay one incentive structure makes subjects behave more risk aversely. However, pay one controls for possible wealth effects compared to pay all, and the authors find evidence that pay one is better under repetition in individual decision making experiments.
List et al. (2005)	Lottery Choice	Using both students and experienced traders, show that while students' choices do not reflect evidence of the reduction of compound lotteries principle, the choices of the professional traders do.
Loomes et al. (1991)	Lottery Choice	Subjects' preferences exhibit cyclical patterns and the direction of the patterns is consistent with regret theory and a violation of transitivity, therefore the cycles are not due to an issue with pay one.
Schmidt and Hewig (2015)	Lottery Choice	Using a within-subjects design, show that subjects make more risky decisions with pay all than with pay one.
Sefton (1992)	Social Preferences	When the probability of payment for each subject is only 25% because only two out of eight pairs are paid, dictator behavior is more generous than when all subjects are paid.
Sherstyuk et al. (2013)	Infinite-horizon Prisoners' dilemma game	Find that pay one causes players to discount the future more heavily than pay all. Additionally, subjects are more myopic and time inconsistent under pay one. Cooperation rates do not differ whether paying for the last round or all rounds but are significantly lower when paying for one round randomly. Subjects use the "Always Defect" strategy significantly more under pay one than under paying for only the last round or paying for all rounds. More subjects use the "Tit-for-Tat" and "Trigger-with-Reversion" strategies under paying for only the last round or paying for all rounds.
Smeets et al. (2015)	Social Preferences	People with over one million Euro in their bank account play dictator game for $100. One out of ten subjects is paid.
Stahl and Haruvy (2006)	Social Preferences	Pay one misrepresents behavior as egalitarian in the form of "path-dependent egalitarian warm glow." Because subjects' choices are only implemented with small probabilities, the cost of an egalitarian option is very low but the benefits of warm glow are felt regardless of the implemented choice. Larger group sizes can minimize other-regarding preferences
Starmer and Sugden (1991)	Lottery Choice	Compare behavior of subjects facing two pairs of lotteries with one randomly selected for payment and subjects facing a single choice. Behavior is consistent across treatments thus there is no evidence that subjects reduce compound lotteries, suggesting that pay one was not problematic.

Option A	Option B
1/10 of \$2.00, 9/10 of \$1.60	1/10 of \$3.85, 9/10 of \$0.10
2/10 of \$2.00, 8/10 of \$1.60	2/10 of \$3.85, 8/10 of \$0.10
3/10 of \$2.00, 7/10 of \$1.60	3/10 of \$3.85, 7/10 of \$0.10
4/10 of \$2.00, 6/10 of \$1.60	4/10 of \$3.85, 6/10 of \$0.10
5/10 of \$2.00, 5/10 of \$1.60	5/10 of \$3.85, 5/10 of \$0.10
6/10 of \$2.00, 4/10 of \$1.60	6/10 of \$3.85, 4/10 of \$0.10
7/10 of \$2.00, 3/10 of \$1.60	7/10 of \$3.85, 3/10 of \$0.10
8/10 of \$2.00, 2/10 of \$1.60	8/10 of \$3.85, 2/10 of \$0.10
9/10 of \$2.00, 1/10 of \$1.60	9/10 of \$3.85, 1/10 of \$0.10
10/10 of \$2.00, 0/10 of \$1.60	10/10 of \$3.85, 0/10 of \$0.10

Fig. 1. Holt and Laury (2002) mechanism.

2.1. Binary-choice lotteries

In the binary-choice lottery paradigm, participants face pairs of lotteries and in each pair they select which lottery they would like to play, and a randomization device then determines the result of the chosen lottery. If a researcher chooses pay all, subjects observe the outcome of each selected lottery and are then paid for each realized outcome. With pay one, though the subjects have submitted decisions for many pairs of lotteries, they are only paid for the outcome of the lottery pair that has been selected randomly. Note that binary-choice-lottery experimental designs already include risk and choosing pay one introduces an additional layer of risk.

Consider a binary-choice-lottery experimental design where subjects face two decisions. Perhaps the most well-known example is the risk-elicitation multiple price list mechanism used in Holt and Laury (2002; see Fig. 1). This involves an array of 10 decision tasks presented in rows and featuring a higher-variance option ("risky") and a lower-variance option ("safe").

The higher-variance option becomes more and more attractive as one descends down the rows, until choosing this option in the bottom row is the dominant strategy. Participants make a choice in each of the 10 rows; one is chosen at random for actual payoff.

The likelihood of the high (low) payoff increases (decreases) as one moves down the Table, so that Option B is increasingly attractive in lower and lower rows; a risk-neutral person will choose Option A in the top four rows and Option B in the bottom six rows. The row in which one switches from choosing Option A to choosing Option B depends on one's risk preferences. In the original study, one row was selected randomly, with the decision made for that row implemented.

2.1.1. Evidence that pay one and pay all lead to similar results

Laury (2006) addresses the concern of diluted incentives with pay one in a multiple-price-list experiment on risk preferences, with the same format as Holt and Laury (2002). In one treatment she pays one of the ten choices at low stakes, while in another all ten choices are paid at low stakes. A third treatment is pay one at high stakes. Laury (2006) finds risk aversion to be highest under high stakes but does not observe a difference in behavior when comparing pay all and pay one under low stakes. Because subjects behave the same in these treatments, there does not appear to be evidence that subjects are scaling down the incentives to account for the random selection.

Starmer and Sugden (1991) present subjects with two pairs of lotteries with one pair randomly chosen for payment, comparing behavior in a single-choice treatment and a pay-one treatment in order to test for whether people reduce compound lotteries to simple ones, as prescribed by expected utility. The authors find evidence supporting the usefulness of pay one. Additionally, Cubitt et al. (1998) find no evidence of cross-task contamination with pay one and no evidence that subjects' choices in pay one exhibit less risk aversion than in the single-choice treatment.

Harrison and Swarthout (2014) show that when implementing pay one, compared to a single-choice experimental design, there is no difference in behavior under the assumption of Expected Utility Theory (EUT). Indeed, Harrison et al. (2013) find that even when faced with multiple pairs of lotteries in pay one, subjects' choices show a significant shift towards risk-neutral behavior from more risk-averse behavior. In contrast, Schmidt and Hewig (2015) show that subjects make more risky decisions with pay all than with pay one, using a within-subjects design. Since expressed risk preferences differ in the two treatments, this creates concern about using pay one. However, within-subject designs and between-subjects designs may lead to different results; see Charness et al. (2013) for more discussion.[4]

In experiments using the binary-choice lottery design where subjects face at least 20 pairs of lotteries, one can observe cycles in subjects' choices. Loomes et al. (1991) investigate if these cycles are the result of a problem with pay one, finding that the direction of the cycles is consistent with regret theory and that they cannot be due to random error. Additionally, they determine that the cycles are the result of a violation of transitivity, rather than reflecting any issue with pay one.

List et al. (2005) use an Allais-paradox experiment to test whether the independence axiom is violated in a study with professional traders from the Chicago Board of Trade. The traders' choices do not offer evidence of difficulties with compound lotteries, so that pay one does not seem problematic in this case. Brokesova et al. (2015) compare behavior on a risk-taking task where it is either the only task and payment are assured, where it is one of several similar tasks of which one will be randomly selected for payment, and whether it is the only task but there is only a small probability of receiving payment. They find no difference in risk-taking behavior when subjects decide between a risky payment and a safe payment across single-choice and pay-one treatments.

2.1.2. Evidence that pay one and pay all lead to different results

There are also studies that show differences in behavior across payment approaches. Cox et al. (2014) presented subjects with pairs of lotteries that include "asymmetrically-dominated" alternatives: An alternative that is inferior in all respects to one option; but, in comparison to the other option, it is inferior in some respects and superior in others (Huber et al., 1982). The authors find that subjects choose the safe option more frequently when listed next to an asymmetrically dominated alternative than when presented in isolation. The preference reversal calls into question the usefulness of pay one in experiments with this design feature. In addition, Cox and Sadiraj (2015) run a thorough and exhaustive test of eight different incentive structures and find large differences in revealed risk preferences across the eight options.

The List et al. (2005) study referenced above also has a treatment involving undergraduate students. In contrast to the results with professional traders, the students do appear to have some difficulty with compound lotteries, creating concern about using pay one.

Nevertheless, our sense is that the overall empirical evidence supports the usefulness of pay one approach in the preponderance of lottery choice experiments, but that concerns arise when the environment deviates from the simple pairwise lottery selection into more complex decisions or when subjects are less sophisticated in terms of complexity and compound lotteries.

[4] Charness et al. (2013) consider the advantages and disadvantages of each type of design. A main point is that within-subject designs may lead to spurious correlations and that between-subjects designs seem more conservative. While between-subject designs lack the statistical power of within-subject designs, significant effects found with this method are more reliably present (Type I error versus Type II error).

2.2. Social preferences

Cox et al. (2008) conducted a test of trust, fear, and reciprocity using both triadic and dyadic designs. Subjects participated in either a single-choice treatment or a pay-one treatment, and there was no difference in subject behavior across the two treatments. In this specific social-preference environment, pay one and single-choice approaches produce similar results.

Stahl and Haruvy (2006) find that pay one distorts behavior in dictator games of varying group sizes, where subjects know the selected dictator prior to making allocation choices, leading them to state "Selecting one decision out of many to determine monetary payoffs can drastically reduce the importance of terminal payoffs relative to path-dependent effects." Clearly, when the dictator is announced after subjects have made their choices and the group size is larger, the probability of one's choice being implemented for actual payment is small. In that situation, subjects appear to behave more pro-socially.

Cox (2009) uses game triads to investigate differences in behavior across social contexts. He finds significant changes in behavior given the strong social context generated by repeated interaction and pay one compared to a single-choice weak social context.

On balance, we do not have a clear recommendation regarding using pay one in social-preference experiments. We find that there is reason for caution, but that there is also evidence that using pay one may be relatively innocuous. Whether the experimenter chooses to implement pay one or pay all in this environment may come down to logistical considerations. For example, Charness and Rabin (2002) considered a large number of simple experimental games and decision tasks in a pen-and-paper experiment. To amass a sufficient number of observations for each game or task, it was necessary to have subjects make choices in a number of different tasks. The authors chose to pay for one of four (or two of eight) decisions, randomly-drawn to ease the process of calculating payoffs and making payments in real time.[5]

2.3. Belief elicitation

Another issue involves incentivizing belief elicitation, an important issue for belief-based models of behavior. Concerns can arise when paying for both the game outcome as well as the incentivized belief elicitation. Subjects may find an opportunity to hedge by, for example, betting against their success in the task. One alternative is to pay only for action choices or beliefs, with the choice of which of these two made randomly. Another is to simply not incentivize beliefs. A third option is to make the payoffs from beliefs relatively small; this approach can substantially reduce incentives to hedge, although of course these are not completely eliminated. A fourth option (used by Armentier and Treich, 2009; Danz et al., 2012) is to pay a subject for a stated belief based on the choice of a subject other than the one with whom she is paired.

Blanco et al. (2010) conduct experiments with the Prisoner's Dilemma and a coordination game, testing whether behavior differs when people are paid for each task or for only one task drawn at random. They find no difference in stated beliefs or choices in the Prisoner's Dilemma, when beliefs are measured using the quadratic-scoring rule. However, they do find a difference in the coordination game using a linear belief-elicitation mechanism. They argue that hedging possibilities are more transparent in their latter case. Overall, their conclusion is that "Hedging can indeed be a problem in belief-elicitation experiments. However – at least according to our results, this seems to be the case only if incentives to hedge are strong and prominent."

So what should an experimenter do when faced with the necessity of eliciting beliefs in a multi-round experiment? The first approach of simply not incentivizing beliefs may be unsatisfactory. Erev et al. (1993) find differences in behavior across incentivized and non-incentivized beliefs in a public-goods game and Croson (2000) also finds differences in a Prisoner's Dilemma. On the other hand, Nyarko and Schotter (2002), Rutström and Wilcox (2009) and Gächter and Renner (2010) do not find evidence of differences in behavior according to whether or not beliefs are incentivized. In any case, we are reluctant to recommend not incentivizing beliefs when this is feasible, given that incentivization has long been considered a strong suit of experimental economics. Regarding paying for either actions or beliefs (but not both), the Blanco et al. (2010) results suggest that this may not be necessary if hedging opportunities are not obvious.

Overall, we favor one of the other two approaches, although we feel that these do need more study. Paying for beliefs about another person's action may not always be feasible, but seems attractive when it can be implemented without confusing the subjects. Alternatively, one could simply make the belief payoffs small relative to the choice payoffs, thereby greatly reducing the scope for hedging. In this case, we do not recommend paying only for either beliefs or choices drawn randomly.

2.4. Auctions

In both first-price sealed-bid private-value auctions (each potential bidder has a separate draw from the range of possible values) and common-value auctions (the value is the same for all bidders), the accumulation of cash balances from earnings significantly affects bidding behavior including evidence of income targeting (Garvin and Kagel, 1994; Ham et al., 2005; Casari et al., 2007). In the beginning, subjects appear to bid high in order to win the auction, but as their cash balance grows,

[5] An additional issue was that the payoffs in each task might have seemed miniscule if subjects were paid for each task. The evidence from Cabrales et al. (2003) mentioned earlier also influenced the choice.

bidding falls as they reach their income target (Ham et al., 2005). This trend in bidding behavior appears more prevalent in inexperienced bidders and diminishes as bidders gain experience in additional sessions (Casari et al., 2007).

Some argue that this issue can be addressed with econometric techniques. For example, in a first-price sealed-bid private value auction, cash balances are endogenous and because only the winner's cash balance adjusts each period, there is little variation in cash balances. Ham et al. (2005) tackle the endogeneity problem in two ways: By adjusting the experimental design to allow for exogenous variation in cash balances, and by using instrumental variables in their estimation of the cash-balance effect based on the new exogenous variation in cash balances, including cash balances as an explanatory variable. Still, after correcting for endogeneity, they find significant evidence of cash balances affecting bidding behavior, as mentioned above.

Overall, our sense is that the concerns about cash-balance accumulation and bankruptcy with pay all support using pay one in auction settings.

2.5. Dynamic-choice and infinite-horizon settings

Comparing behavior in a pay-one treatment and a single-choice treatment in a dynamic-choice setting ("Deal or No Deal"), Baltussen et al. (2012) find, on average, that risk preferences were consistent across the single-choice and pay-one treatments. This result supports the approach of paying each subject for one of her 10 decisions in dynamic-choice experiments, allowing the researcher to obtain more data and make within-subject comparisons.

Sherstyuk et al. (2013) assess behavior in an infinite-horizon prisoner's-dilemma game; the infinite horizon is implemented using a stochastic termination rule. Paying subjects for a randomly-selected round as opposed to all rounds resulted in lower cooperation rates, more myopic behavior, and a present period bias. Subjects showed a higher probability of adopting the "Always Defect" strategy rather than the "Tit-For-Tat" as found with pay all. While paying all subjects in an infinitely-repeated game can get costly, especially when the incentives must be large enough to ensure subject motivation, it appears that *paying for one randomly selected round* changes behavior relative to pay all.

Sherstyuk et al. (2013) also investigate if paying for only the last round might provide similar results to paying for all rounds in the infinite-horizon environment. The authors find the same cooperation rates and the same strategy profiles in these two treatments. Since paying for only the last round avoids wealth and portfolio effects, this approach (with infinite-horizon environments) seems useful; Chandrasekhar and Xandri (2014) develop theory indicating that this is the best approach.

3. Paying only a subset of the participants: evidence

In some experiments, such as classroom or online, paying only a subset of the participants is desirable from a logistics perspective. Paying only one out of many participants combines the issues discussed above in relation to making the actual pay for a given action probabilistic. In addition, there is potentially a fundamental psychological difference between being definitely paid a positive amount and only having a chance to receive payment, suggesting caution concerned paying a random subset of the experimental subjects.

In general, the results in the literature are encouraging because they show little difference between the two methods. In an early experiment comparing the pay-all to the pay-subset mechanism, Bolle (1990) conducted an ultimatum game experiment comparing the pay-all treatment to a treatment in which only two out of twenty subjects were paid. There was no significant difference in behavior between the treatments.

Sefton (1992) tested the effect of paying all participants versus paying 2 out of 8 pairs in a $5 dictator game. He reports that participants are more generous when only one in four is paid. Note that the dictator game is particularly sensitive to the actual amount of payment (Forsythe et al., 1994) compared for example with payments in the ultimatum game (Slonim and Roth, 1998; Andersen et al., 2011). In contrast with Sefton (1992), Clot et al. (2015) compare the results of a dictator game treatment in which all participants were paid with one in which only a subset was paid, and report no significant difference between the treatments.

Beaud and Willinger (2015) conduct risk-elicitation experiments, using an adaptation of the Gneezy and Potters (1997) risk-elicitation method, while varying different aspects of methodology. In Experiment 1, participants could earn up to 250 Euro, but only 10% of the participants were randomly selected at the end of the experiment to actually be paid. In Experiment 2, all participants were paid according to their earnings in the experiment, but stakes were much lower than in experiment 1 (the maximum was 50 Euro). The main finding is that there is no difference in risk vulnerability (reflecting sensitivity to background risk) across these two designs.[6] This was true despite of the fact that many other features were varied across these treatments.[7] Brokesova et al. (2015) tested how a small probability of receiving payment affected subjects' risk-taking

[6] A related potential concern with pay one is the introduction of "background risk," described as any risk that is "committed but unresolved" (see also Lee, 2008). A randomization device that determines subjects' payments after all decisions are made and is not resolved before subjects make their decisions introduces this issue. If the introduction of such background risk alters subjects' behavior, pay one is a less attractive approach.

[7] Experiment 1 was a pen-and-paper experiment, while Experiment 2 was computerized. Participants in Experiment 1 were first-year master's students in economics, while participants in Experiment 2 were selected from a subject pool from various disciplines. Finally participants in Experiment 1 received an endowment from the experimenter, while participants in Experiment 2 had to "work" in a preliminary task to earn a monetary endowment.

behavior compared to a single-choice treatment. In the experiment, subjects faced a choice between a safe and a risky payment that mirrored the structure of a local bank promotion. Again, they find no difference in subjects' choices.

Harrison et al. (2007) tested the effect of increasing the payoffs in the Holt and Laury (2002) design discussed above, using payoffs that are roughly 150 times the base payoffs in Holt and Laury (2002), but paid only 10% of their participants. They report an additional experiment that directly examine the hypothesis that paying only a subset of the participants generates the same responses as paying all of them the large amount. Harrison et al. (2007) report that they could not reject the null hypothesis that paying all the participants and paying only 10% of them results in the same behavior. Similarly, Andersen et al. (2014) explicitly tested the effect of paying only a subset of the participants by exogenously varying the probability of payment for their discounting task from 10% to 100%. They conclude: “the effect of probabilistic discounting is non-existent or negligible in our sample, and for the specifications considered here.”

Baltussen et al. (2012) is an exception to the above, demonstrating concerns about paying a random subject, particularly in a complex and dynamic environment. They use the game show “Deal or No Deal” to test this type of dynamic-choice environment. In the experimental design, they used a baseline treatment where subjects played the game once for payment. As discussed earlier, one treatment of interest included subjects playing 10 rounds with one of these rounds randomly selected for payment. In another treatment subjects played a single round and 10% of subjects were randomly selected to receive payment.[8] While the paying-a-random-subject treatment and baseline treatment avoid any cross-task contamination concerns, the authors observe a significant decrease in risk aversion in the random-subject treatment.

Overall, the majority of comparisons of paying all the participants versus only a subset of them indicate that the loss of motivation is small—much smaller than the implied reduction in actual payment.

An important issue relates to moral choices in cases where pay only one out of N participants is used. In many experiments, the decision made by the participant includes some moral value. In such situations, participants may care about signaling to others that they are moral people. They might also be concerned about identity, and hence would use their choices to self-signal how moral they are. In situations where signaling is important, paying only some of the participants may be problematic because the moral act is chosen with certainty, while the material payoff is paid only in a fraction of the cases. This problem could affect moral and immoral choices.

Consider for example Smeets et al. (2015), who conduct a dictator game with people that have more than one million Euro in a bank account. These participants receive 100 Euro, and are asked whether they wish to give some of this money to another participant. However, only one out of ten participants is paid according to the decision he/she made. If a participant decides to send a large sum of money, say the entire 100 Euro, then he/she can feel good about the action, and gain the value of positive signaling. The positive signaling value from sending 100 Euro could be large. Yet, the expected cost of this decision is only 10 Euros, because only one out of ten participants is being paid. For an early example of studying this issue, see Stahl and Haruvy (2006).

A similar case, but with a negative action, relates to lying. Participants in Erat and Gneezy (2012) play a deception game in which one party sends a message advising the second party about which action pays the receiver the most. If the advice given is followed, in most cases lying increases the payoff of the sender but decreases the payoff of the receiver, while in some cases lying increases the payoffs for *both* the sender and receiver. In the experiment, only one out of twenty senders was paid, and the conclusion was that some people prefer not to lie even when lying increases theirs and the receiver's payoff. However, as in the dictator game, this conclusion might be inflated because a sender who chooses to lie is a liar—regardless of whether he or she subsequently receives payment. The signaling value is certain, while the monetary payoff is not.

Future research could investigate whether this difference between the signal value and its cost is indeed larger when only some of the participants are actually paid. If it is, then the conclusion drawn from these studies is inflated in the direction of considering people to be moral. If the difference is not important, then this is a very efficient way of collecting data.

4. Conclusion

Experimental research in economics differs from many disciplines in that researchers generally incentivize subjects' decisions. When selecting an incentive structure, researchers need to assess various characteristics including risk of cross-task contamination, bankruptcy, income effects, and data analysis concerns. While designing an experiment around a single choice is the cleanest approach, such experiments can be expensive and only allow for between-subject data analyses. Gathering data from multiple periods is considerably more efficient than a single-choice design. In addition, more complex decisions may require multiple periods to ensure comprehension.

Yet having subjects complete multiple decisions and paying for each one comes with its own set of issues including the potential for wealth and portfolio effects, cross-task contamination, hedging opportunities, and bankruptcy issues. One alternative approach that we find is generally useful is to have subjects make multiple decisions and pay for a randomly-selected one. But using pay one raises its own concerns: while this eliminates problems associated with wealth and portfolio effects or bankruptcy, there is the potential for diluted incentives and the introduction of background risk. Another technique, which minimizes transactions costs, is to be pay only a subset of the participants. The evidence we have presented suggests that there may not be a substantial loss of motivation when this device is used.

[8] The face value of prizes was held constant across treatments (guaranteed payment, random round, and random subject).

Table 2
Summary of recommendations.

Environment	Recommendation
Lottery Choice/Multiple Price Lists	Empirical evidence supports the usefulness pay one in the great majority of lottery choice experiments but concerns arise when the environment deviates from the simple pairwise lottery selection into more complex decisions.
Social Preferences	Because social preferences span such a variety of environments and the evidence is mixed, we do not make a strong recommendation. It is not clear that pay one leads to different results, so in principle either approach could be employed. Logistical issues may affect the decision to use pay one or pay all.
Belief Elicitation	If hedging opportunities are not obvious, pay one is a useful method. A more cautious approach is to pay for both the game outcome and the belief elicitation but to make the belief payoffs small relative to the choice payoffs, thereby greatly reducing the scope for hedging.
Auctions	The concerns about cash balance accumulation and bankruptcy with a pay all incentive structure support pay one over pay all in auction settings.
Background Risk	Pay one is not problematic if the background risk is resolved prior to the subject making any decisions. For example, handing the subject a sealed envelope containing the randomly-selected round for payment.
Dynamic Settings	Only one study of which we are aware. Pay one and pay all lead to similar behavior in a "Deal or No Deal" setting and so we endorse pay one here. However, paying only 10% of subjects led to a significant decrease in risk aversion, so we do not recommend this approach in this setting. More research is needed.
Infinite Horizon	Only one study of which we are aware. Shown both theoretically and empirically in infinite-horizon settings with random termination, subject behavior is consistent whether paying for all rounds or when paying for the last round. More research is needed.

Table 2 summarizes our recommendations in the different types of experiments that we have discussed.

We have detailed existing evidence for and against each incentive structure, providing examples of experiments comparing subjects' behavior under different incentive structures. In general, while in some cases the pay-one (and pay-a-subset) method may distort behavior, the data suggest that in the majority of cases it is either equal to (or sometimes superior to) the pay-all method. In general, we feel that both pay one and pay all are useful methods, although there are instances (such as when bankruptcy is possible) where practical considerations suggest using pay one (as is indicated by the theoretical work of Azrieli et al., 2015). We hope to further the discussion about how to best choose an incentive structure when designing an experiment.

References

Andersen, S., Ertac, S., Gneezy, U., Hoffman, M., List, J., 2011. Stakes matter in ultimatum games. Am. Econ. Rev. 101, 3427–3439.
Andersen, S., Harrison, G., Lau, M., Rutström, E., 2014. Discounting behavior: a reconsideration. Eur. Econ. Rev. 71, 15–33.
Armentier, O., Treich, N., 2009. Subjective probabilities in games: an application to the overbidding puzzle. Int. Econ. Rev. 50, 1079–1102.
Azrieli, Y., Chambers, C., Healy, P., 2015. Incentives in Experiments: A Theoretical Analysis (Unpublished manuscript).
Baillon, A., Halevy, Y., Li, C., 2015. Experimental Elicitation of Ambiguity Attitude Using the Random Incentive System (Unpublished manuscript).
Baltussen, G., Post, G., van den Assem, M., Wakker, P., 2012. Random incentive systems in a dynamic-choice environment. Exp. Econ. 15, 418–443.
Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., Sugden, R., 2009. Experimental Economics: Rethinking the Rules. Princeton University Press, Princeton.
Beaud, M., Willinger, M., 2015. Are people risk vulnerable? Manage. Sci. 61, 624–636.
Blanco, M., Engelmann, D., Koch Normann, A.H.-T., 2010. Belief elicitation in experiments: is there a hedging problem? Exp. Econ. 13, 412–438.
Bolle, F., 1990. High reward experiments without high expenditure for the experimenter? J. Econ. Psychol. 11, 157–167.
Brokesova, Z., Deck, C., Peliova, J., 2015. Bringing a Natural Experiment into the Laboratory: the Measurement of Individual Risk Attitudes (Unpublished manuscript).
Cabrales, A., Charness, G., Corchón, L., 2003. An experiment on nash implementation. J. Econ. Behav. Organiz. 51, 161–193.
Casari, M., Ham, J., Kagel, J., 2007. Selection bias, demographic effects, and ability effects in common value auction experiments. Am. Econ. Rev. 97, 1278–1304.
Chandrasekhar, A., Xandri, J., 2014. A Note on Payments in the Lab for Infinite Horizon Dynamic Games with Discounting (Unpublished manuscript).
Charness, G., Levin, D., 2009. The origin of the winner's curse: a laboratory study. Am. Econ. J. Microbiol. 1, 207–236.
Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Quartly J. Econ. 117, 817–869.
Charness, G., Gneezy, U., Imas, A., 2013. Experimental methods: eliciting risk preferences. J. Econ. Behav. Organiz. 87, 43–51.
Clot, S., Grolleau, G., Ibanez, L., 2015. Shall We Pay All? An Experimental Test of Random Incentivized Systems (Unpublished manuscript).
Cox, J., Sadiraj, V., Schmidt, U., 2015. Paradoxes and mechanisms for choice under risk. Exp. Econ. 18, 215–240.
Cox, J., Sadiraj, K., Sadiraj, V., 2008. Implications of trust, fear, and reciprocity for modeling economic behavior. Exp. Econ. 11, 1–24.
Cox, J., Sadiraj, V., Schmidt, U., 2014. Alternative payoff mechanisms for choice under risk. Int. Adv. Econ. Res. 20, 239–240.
Cox, J., 2009. Trust and reciprocity: implications of game triads and social contexts. N. Z. Econ. Pap. 43, 89–104.
Croson, R., 2000. Thinking like a game theorist: factors affecting the frequency of equilibrium play. J. Econ. Behav. Organiz. 41, 299–314.
Cubitt, R., Starmer, C., Sugden, R., 1998. On the validity of the random lottery incentive system. Exp. Econ. 1, 115–131.
Cubitt, R., Starmer, C., Sugden, R., 2001. Discovered preferences and the experimental evidence of violations of expected utility theory. J. Econ. Methodol. 8, 385–414.
Danz, D., Fehr, D., Kübler, D., 2012. Information and beliefs in a repeated normal-form game. Exp. Econ. 15, 622–640.
Erat, S., Gneezy, U., 2012. White lies. Manage. Sci. 58, 723–733.
Erev, I., Bornstein, G., Wallstein, T.S., 1993. The negative effect of probability assessments on decision quality. Organiz. Behav. Human Dec. Proc. 55, 78–94.

Forsythe, R., Horowitz, J., Savin, N., Sefton, M., 1994. Fairness in simple bargaining experiments. Games Econ. Behav. 6, 347–369.
Gächter, S., Renner, E., 2010. The effects of (incentivized) belief elicitation in public goods experiments. Exp. Econ. 13, 364–377.
Garvin, S., Kagel, J., 1994. Learning in common value auctions: some initial observations. J. Econ. Behav. Organiz. 25, 351–372.
Gneezy, U., Potters, J., 1997. An experiment on risk taking and evaluation periods. Q. J. Econ. 112, 631–645.
Ham, J., Kagel, J., Lehrer, S., 2005. Randomization, endogeneity, and laboratory experiments: the role of cash balances in private value auctions. J. Econom. 125, 175–205.
Harrison, G., Swarthout, T., 2014. Experimental payment protocols and the Bipolar Behaviorist. Theory Decis. 77, 423–438.
Harrison, G., Lau, M., Rutström, E., 2007. Estimating risk attitudes in Denmark: a field experiment. Scand. J. Econ. 109, 341–368.
Harrison, G., Martinez-Correa, J., Swarthout, T., 2013. Inducing risk neutral preferences with binary lotteries: a reconsideration. J. Econ. Behav. Organiz. 94, 145–159.
Holt, C., Laury, S., 2002. Risk aversion and incentive effects. Am. Econ. Rev. 92, 1644–1655.
Holt, C., 1986. Preference reversals and the independence axiom. Am. Econ. Rev. 76, 508–515.
Huber, J., Payne, J., Puto, C., 1982. Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. J. Consum. Res. 9, 90–98.
Karni, E., Safra, Z., 1987. 'Preference reversal' and the observability of preferences by experimental methods. Econometrica 55, 675–685.
Laury, S., 2006. Pay One or Pay All: Random Selection of One Choice for Payment (Unpublished manuscript).
Lee, J., 2008. The effect of the background risk in a simple chance improving decision model. J. Risk Uncertainty 36, 19–41.
List, J., Haigh, M., Nerlove, M., 2005. A simple test of expected utility theory using professional traders. Proc. Natl. Acad. Sci. 102, 945–948.
Loomes, G., Starmer, C., Sugden, R., 1991. Observing violations of transitivity by experimental methods. Econometrica 59, 425–439.
Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs. Econometrica 70, 971–1005.
Rutström, E., Wilcox, N., 2009. Stated beliefs versus inferred beliefs: a methodological inquiry and experimental test. Games Econ. Behav. 67, 616–632.
Schmidt, B., Hewig, J., 2015. Paying out one or all trials: a behavioral economic evaluation of payment methods in a prototypical risky decision study. Psychol. Rec. 65, 245–250.
Sefton, M., 1992. Incentives in simple bargaining games. J. Econ. Psychol. 13, 263–276.
Sherstyuk, K., Tarui, N., Saijo, T., 2013. Payment schemes in infinite-horizon experimental games. Exp. Econ. 16, 125–153.
Slonim, R., Roth, A., 1998. Learning in high stakes ultimatum games: an experiment in the Slovak Republic. Econometrica 66, 569–596.
Smeets, P., Bauer, R., Gneezy, U., 2015. Giving behavior of millionaires. Proc. Natl. Acad. Sci. 34, 10641–10644.
Stahl, D., Haruvy, E., 2006. Other-regarding preferences: egalitarian warm glow, empathy, and group size. J. Econ. Behav. Organiz. 61, 20–41.
Starmer, C., Sugden, R., 1991. Does the random-lottery incentive system elicit true preferences? an experimental investigation. Am. Econ. Rev. 81, 971–978.