CBESS Discussion Paper 15-20R2

Mathematics self-confidence and the "prepayment effect" in riskless choices

by Lian Xue*
Stefania Sitzia*
Theodore L. Turocy*

*School of Economics and CBESS, University of East Anglia

Abstract

We extend the analysis of a riskless choice experiment reported recently by Hochman et al. (2014). Participants select from among sets of standard playing cards valued by a simple formula. In some sessions, participants are given a prepayment associated with some of the cards, which need not be the earnings-maximizing ones. Hochman et al. find that participants choose an earnings-maximizing card less frequently when another card is prepaid. We replicate this result under the original instructions, but not with instructions which explain the payment process more explicitly. Participants who state they do not consider themselves good at mathematics make earnings-maximizing choices much less frequently. Prior experience with economics experiments in general improves performance only modestly. The results suggest that even when comparisons among choices require only simple quantitative reasoning steps, market designers and regulators may need to pay close attention to how the terms of offers are expressed, explained, and implemented.

JEL classification codes
C91; D83

Keywords
loss aversion; prepayment; replication; mathematics self-confidence; lab rats.

Centre for Behavioural and Experimental Social Science
University of East Anglia
Norwich Research Park
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/cbess

Mathematics self-confidence and the "prepayment effect" in riskless choices*

Lian Xue Stefania Sitzia Theodore L. Turocy†

School of Economics and
Centre for Behavioural and Experimental Social Science
University of East Anglia
Norwich NR4 7TJ
United Kingdom

January 23, 2017

Abstract

We extend the analysis of a riskless choice experiment reported recently by Hochman et al. (2014). Participants select from among sets of standard playing cards valued by a simple formula. In some sessions, participants are given a prepayment associated with some of the cards, which need not be the earnings-maximizing ones. Hochman et al. find that participants choose an earnings-maximizing card less frequently when another card is prepaid. We replicate this result under the original instructions, but not with instructions which explain the payment process more explicitly. Participants who state they do not consider themselves good at mathematics make earnings-maximizing choices much less frequently overall, but those who express self-confidence in mathematics drive the treatment effect. The results suggest that even when comparisons among choices require only simple quantitative reasoning steps, market designers and regulators may need to pay close attention to how the terms of offers are expressed, explained, and implemented.

JEL Classifications: C91; D83

Keywords: loss aversion; prepayment; replication; mathematics self-confidence.

*We thank Guy Hochman for sharing the data from the original experiment, and Piers Fleming, Robert Sugden, and seminar participants at University of East Anglia for comments and suggestions, as well as the helpful comments of two anonymous referees and the associate editor. Turocy acknowledges the support of the Network for Integrated Behavioural Science (Economic and Social Research Council Grant ES/K002201/1). All errors are the sole responsibility of the authors.

†Corresponding author: `T.Turocy@uea.ac.uk`

1 Introduction

A stylized fact of behavioral economics is that individuals dislike losses more than they like equivalent gains. The endowment effect (e.g. Thaler, 1980; Kahneman et al., 1990), status quo bias (e.g. Kahneman et al., 1991), and the sunk cost fallacy (e.g. Thaler, 1999) are well-established behavioral regularities which can be attributed to loss aversion. Some studies of these effects involve cross-modal comparisons involving different types of goods or assets. For example, Kahneman et al. (1990) observed that participants were reluctant to sell a mug given to them by the experimenter at the beginning of the experiment. Evidence is mixed on whether the endowment effect extends to money, as opposed to other goods.[1] Becker et al. (1974) found that participants in a non-incentivized experiment systematically treated outlay costs differently from opportunity costs. However, Kahneman et al. did not find evidence of the effect when they used tokens instead of mugs.

Hochman et al. (2014) (henceforth HAA) report an experiment in which participants perform a series of choice tasks with no objective risk. In each task, participants are presented with a set of four standard playing cards, from which they must choose one. To each card is attached a monetary value, which depends on both the suit and the rank of the card. In one treatment, participants receive an advance payment, which corresponds to a specific pattern of choices across five out of seven decision tasks. Some of these choices are not the optimal (earnings-maximizing) ones for their corresponding task. In a control treatment, participants receive no advance payment. They find that participants who receive prepayment are more likely to carry out the pre-paid choice even when it is not earnings-maximizing, while participants who do not receive prepayment are relatively more likely to make the earnings-maximizing choice. As the choice task involves no objective risk, HAA interpret this result as supporting loss aversion in money in a very strong form, and label this the "prepayment effect."

Such a strong result could have significant implications for the choice architectures firms use to interact with employees, contractors, and consumers. HAA give an example in which a firm pre-pays its downstream agents their commissions for selling their product, with the requirement that those agents pay back commissions should they fall short of their target; the prepayment effect would suggest these agents are more likely to stick with selling the firm's product (as opposed to a rival's) when they are paid in advance. The phenomenon of sticking to an apparently suboptimal decision has a parallel in consumer decision-making, in which it is argued that consumers fail to switch suppliers in markets such as banking, energy, and telecommunications, even though better plans are on offer. (Wilson and Waddams-Price, 2010; Lunn, 2011; Grubb, 2015; Grubb and Osborne, 2015; Sitzia et al., 2015)

[1] For a lengthier discussion see Hochman et al. (2014).

In those markets, complexity has been cited as one driver of stickiness in both empirical and laboratory studies. (Larrick and Soll, 2008; Allcott, 2011; Kalayci and Potters, 2011; Sitzia and Zizzo, 2011; Grubb and Osborne, 2015) HAA's card-choice task has only one complicating feature: the rule to calculate the value of a card depends on its suit. Spades are worth USD 0.25 times the rank of the card, while all other suits are worth USD 0.10 times the rank of the card. The values of the cards are thus not determined by a single, common standard (Gaudeul and Sugden, 2012; Piccione and Spiegler, 2012) but depend on two dimensions, suit and rank. In two of the three conditions reported by HAA, participants are provided with the monetary value of the cards, either instead of or in addition to the card's suit and rank. In those conditions, in which the values of the cards have been in effect reduced to a common standard, HAA do not observe a prepayment effect. HAA's experiment demonstrates that the introduction of a minimally non-common standard can be enough to affect decision-making significantly.

In this paper, we take a closer look at HAA's card-choice task to understand more completely the underpinnings of the prepayment effect. In Section 2, we conduct an analysis of the experimental design, focusing on their "distant representation" treatment in which values of cards are not presented using a common standard. Based on previous literature, we identify a number of candidate factors which might affect a participant's propensity to choose a card which does not maximize their earnings. These include features of the choice architecture – specifically, the content and structure of the experimental instructions – as well as demographic factors including the participant's prior experience in experiments and their self-perception of their ability in mathematics.

Our results in Section 3 show that behavior in this simple task depends on these factors. We replicate the prepayment effect reported by HAA using their experimental protocol and instructions; prepayment can indeed lower the frequency of earnings-maximizing choices. We show there are significant and substantial differences in the probability of earnings maximization as a function of individual characteristics. Participants who report they are good at mathematics maximize far more often overall, except, notably, in the prepayment treatment as implemented by HAA. It is this group of math-confident participants who account for most of the treatment effect in our data. We find no significant effect due to experience with economics experiments. We also find that the prepayment effect largely disappears under rewritten instructions which describe the transactions that will take place should the participant choose a card other than the one for which prepayment was made. We conclude in Section 4 by discussing how the results from this simple task, which turn out to be quite rich and interesting, might inform the applicability of the prepayment effect beyond the laboratory.

2 Experimental analysis and design

Our experiment replicates and extends the “distant representation” condition of Experiment 1A in HAA. There are seven decision tasks. In each task, a participant sees a set of four playing cards, from which one must be chosen. Each card’s monetary value depends both on its suit and its rank, where aces are considered to have rank one, and jacks, queens, and kings to have ranks 11, 12, and 13, respectively. There are 5 spades cards in the deck, from ace to five; these cards are worth GBP 0.25 times the card’s rank. The value of a card from any other suit is GBP 0.10 times its rank. In HAA spades are valued at USD 0.25 per rank and other suits at USD 0.10 per rank; our stakes were 50 to 60 percent higher than HAA as measured by using the prevailing exchange rate at the time of the experiment.

The seven sets of cards, which are provided in Appendix A, are presented in the same order for each participant. In two sets, a spade is the earnings-maximizing card; these are called “high-spades” trials. In two other sets, a spade is present, but there is some other non-spade card in the set which is worth strictly more; these are called “low-spades” trials. HAA call the remaining three sets filler trials: of these, one set has a spade and a non-spade card which are tied as earnings-maximizing cards, and the remaining two sets do not contain any spade cards. Participants make their choices for all seven sets of cards without any feedback; a summary screen appears at the end of the session reporting the seven chosen cards, and calculating the participant’s earnings for the session.[2]

The experiments were conducted in the laboratory of the Centre for Behavioural and Experimental Social Science (CBESS) at the University of East Anglia. Participants were recruited from the lab’s standard subject pool, which is managed using the hRoot system (Bock et al., 2012). The choice task was computerized using zTree (Fischbacher, 2007). Sessions took place between November 2014 and October 2015, and lasted around 20 to 30 minutes, including all instructions and final payment.

2.1 Baseline and replication

Our first objective is to replicate the existence of the prepayment effect within the subject pool at our laboratory. We conduct two treatments, **Post** and **Pre+Amt**, which replicate the two payment conditions used by HAA.[3] For these treatments, we used instructions which deviated as little as possible from HAA’s, allowing only for the different currency. Treatment **Post** is a standard post-payment setup: participants make all seven choices, then see a summary screen which recaps

[2]Full instructions and screenshots of the decision interface are available as an online appendix.

[3]We present treatment names in boldface type, choosing names intended to help recall the purpose of each treatment.

the decisions and tabulates their earnings. They are then dismissed one-by-one to the payment station and receive their payments in private before departing the laboratory. Treatment **Pre+Amt** implements HAA's prepayment treatment. Each participant is given GBP 3.75 in cash at their station prior to making their choices. Participants are told this amount is the sum value of the five spades cards which will appear among the seven sets (hence the "Amount" in the treatment's label). After all seven choices are made, the summary screen appears; this screen tabulates separately the values of the cards chosen, as well as the value of any spade cards not chosen. Participants are then dismissed one-by-one to the payment station to settle payments in private. We carefully kept to the prepayment framing, in that, for any spades not chosen, we took back coins from the participant, and then gave them different coins for the cards actually chosen; that is, we did not integrate the two payments into a net payment.[4]

Hypothesis 1. *Participants will choose the earnings-maximizing card more often in **Post** than **Pre+Amt**, replicating the existence of the prepayment effect.*

2.2 Influence of the choice environment

HAA attribute the treatment effect they observe between **Post** and **Pre+Amt** to the existence of the prepayment in **Pre+Amt**. We identify two other ways in which the description of the experimental task differs between the two treatments. Exploring alternative instructions serves to map out the robustness of the treatment effect in the lab, but also provides a link to possible applicability in markets. The description of the experimental task is the parallel of the contract between firm and downstream sales agents in HAA's example, or the terms of a tariff in a consumer's plan for electricity, gas, or telecommunications.

In **Pre+Amt**, in order to explain the significance of the GBP 3.75 received in advance of decisions, the instructions included the following language:

> We have placed the cards ace through five of spades in the deck randomly ... The value of the five spades (ace through five) equals a total of three pounds and seventy-five cents: one plus two plus three plus four plus five is fifteen times 25p i.e. three pounds and seventy five pence. We will give you this amount up front.

No parallel language exists in **Post**; the total value of spades is not pointed out to participants explicitly. To identify whether the mere mention of the value of spades might call attention to spades, or otherwise create a reference point, we conduct treatment **Post+Amt**, in which we retain postpayment, but include the language mentioning the total value of spades.

[4] As all decisions were made by this point in the session, this bit of theatre could not affect results; nevertheless we wanted to be careful to maintain the framing throughout the session.

Hypothesis 2. *Participants will maximize less frequently in* ***Post+Amt*** *than in* ***Post****; mentioning the value of spades will result in fewer maximizing choices.*

A second difference between **Post** and **Pre+Amt** is the length and complexity of the instructions. Describing the prepayment protocol necessarily makes instructions longer, as there are more steps to explain to participants.

To check on the robustness of the prepayment effect to alternate but in-principle equivalent ways of explaining the mechanism, we replace some of the text from the HAA originals. Specifically, at the end of their instructions, HAA state,

> At the end of the game, if you have not selected all spades, we will pay you for the cards you have selected, and you will refund us money for the spades you have not selected.

Our design of alternative phrasing is based on HAA reporting of results from follow-on experiments, which they interpret as indicating that an explicit linkage between amounts of money and choices of particular cards is important for the prepayment effect to operate. In our treatment **Pre+Amt+Instr**, we therefore replace the above sentence with the following text intended to highlight that linkage:

> For each set of four playing cards, there are three possible scenarios:
>
> - There is not a spade among the four cards. Then, at the end of the session, you will receive a payment for that set equal to the value of the card you select.
> - There is a spade among the four cards, but you select a different card. Then, at the end of the session, you will pay us back from your up front payment an amount equal to the value of the spade, and you will receive a payment from us equal to the value of the card you did select.
> - There is a spade among the four cards, and you select the spade. Then, because you have already received payment for that spade card in your up front payment, you will not pay us back anything for that set, nor will you receive any additional payment for that set.

Hypothesis 3. *The effect of prepayment depends on the instructions; maximization rates will be higher in* ***Pre+Amt+Instr*** *than* ***Pre+Amt****.*

	Information on total value of spades	Advance payment	Instructions same as HAA
Post	No	No	Yes
Pre+Amt	Yes	Yes	Yes
Post+Amt	Yes	No	No
Pre+Amt+Instr	Yes	Yes	No

Table 1: Summary of experimental treatments.

2.3 Effects of prior experience

We now turn our attention to the characteristics of the decision-makers themselves. Our next hypothesis concerns the role of experience. The card choice task in HAA is simple. It is also artificial, in the sense that the rules for determining valuations of the cards are not drawn from a game or situation that exists "in the wild." Participants cannot draw on prior experience with the task or very similar ones, in the way that an economic agent might draw on experience in understanding the implications of prepayment arrangements for commissions.

The artificiality of the task is a useful feature, in that it allows the construction of a decision problem with the most minimal amount of complexity arising from the lack of a common standard.[5] While we cannot easily address the role that familiarity with this specific task might play in the persistence of the prepayment effect, we can ask whether participants who have previous experience with economic experiments behave differently.

Previous experience might predict behavior for two reasons. First, previous experience may affect expectations about how lab experiments operate in a generalized sense. Experience has been shown to have an effect in, for example, public goods games (Conte et al., 2014) and allocation games (Matthey and Regner, 2013). Some of the effect of experience in games would be due to more accurate beliefs about the play of others, an aspect that is absent in this individual decision task. Nevertheless prior experience may be valuable, insofar as participants will have familiarity with general lab procedures, including how to read and extract relevant information from instructions. Second, participants with prior experience also have self-selected into coming back to the laboratory. Abeler and Nosenzo (2014) have studied self-selection into participant pools and report that interest in monetary rewards appears strongly to drive participation.

Both channels would suggest experienced participants would be more likely to maximize earnings. To explore the possible role of experience with experiments, we conducted a stratified recruiting strategy. We identified very experienced participants as those who had participated at least 10 times previously in experimental sessions, and somewhat less experienced participants as those

[5] Schram (2005) is a good discussion of the tradeoffs between internal and external validity arising from the artificiality of laboratory games.

who had participated no more than 5 times. We recruited approximately equal numbers of participants from these two subpopulations. Those who had participated 6 to 9 times were not recruited, to give a more clear distinction between the groups.

Hypothesis 4. *Participants with greater experience in experiments will have higher maximization rates in prepayment treatments.*

2.4 Effects of confidence at the specific task

The distant representation condition we focus on in our experiment is one of three conditions in the full Experiment 1A design of HAA. In their close representation condition, payoffs were represented not by playing cards, but directly expressed as amounts of money. In the moderate representation condition, payoffs were depicted using playing cards, but with a label indicating their monetary value. HAA find evidence of the prepayment effect only in the distant representation condition.

The distant representation requires a particular type of operation, mathematical calculation, for participants to infer the earnings consequences of the options they face. The arithmetic required is taught at an early age in schools, and can be taken as part of the assumed skills an undergraduate student at a university would have. Although the arithmetic is straightforward enough, studies such as Pajares and Miller (1994) have noted that many adults dislike and avoid math, even those who are competent at calculation. Ashcraft (2002) observes effects of math anxiety even on simple whole-number arithmetic problems. It has been argued that math anxiety could shape individuals' behavior when facing challenging circumstances (Bandura, 1977).

The design of the card choice task requires participants to carry out, on demand, a mathematical calculation in the context of a novel setting not previously encountered. Prepayment may be an effective tactic when the task to determine the earnings-maximizing choice requires an activity the decision-maker is disinclined to carry out. At the end of the session, after all choices were made but before the results of the decisions were shown, participants completed a screen of demographics questions, asking for the participant's gender, current course of study, and whether they were a native speaker of English. In addition, we included the question,, "Do you consider yourself good at mathematics?" This was implemented as a radio box, with options for Yes, No, or "Prefer not to say" as the possible responses. Such self-reports of mathematical skill have been used as a behavioral indicator in economics and psychology (e.g. Ashcraft and Kirk, 2001; Ashcraft and Ridley, 2005; Marsh et al., 2012; Buser et al., 2014).[6]

[6] We prefer the participant's self-reported confidence level rather than their actual mathematical ability. We placed no time restrictions or pressures on the individual decisions, and we are confident our students are capable, in principle at least, of carrying out the required calculations correctly. Our hypothesis instead centers around the possibility that some participants feel less attracted or inclined to engage with the required computational task.

Hypothesis 5. *Participants who report confidence in their ability to carry out mathematical calculations will have higher maximization rates in prepayment treatments.*

3 Results

We report on 206 participants who participated in the task. Within our sample, the breakdown of males (44.2%) and females[7] is comparable to our participant pool and the University's student body as a whole. Less-experienced participants had participated, on average, in 1.78 previous sessions, while the very-experienced participants had been in 15.67 previous sessions on average. A total of 106 (51.5%) considered themselves good at mathematics, with 86 (41.7%) saying they were not; 14 preferred not to say.

We conducted Fisher's exact test for independence between each pair out of the four demographic characteristics of math-confidence, prior experience, gender, and native English speaking.[8] The only pair for which the null hypothesis of independence is rejected at the 10% level is between gender and math-confidence: 61.5% of males (56 in total) answered this question in the affirmative, as opposed to 43.9% of females (50 in total); the difference is significant with a p-value of 0.008. Previous studies (Eccles, 1998; Buser et al., 2014) have also observed greater reported confidence in mathematics among males than females.

The results in HAA focus on the rate of maximization by participants in the two "low-spades" trials, in which a spade card is present but is not the earnings-maximizing choice. We report this measure for each treatment in Table 2, where we also present breakouts of the low-spade maximization rate overall for each characteristic. The rightmost three columns in Table 2 give the percentages of participants who maximize on none, one, or both of the low-spades trials.[9] In general, the modal outcome for a participant is to maximize on both low-spades trials, with maximizing on exactly one of the two being more likely than failing to maximize on either.

We will base our tests principally on the distributions of the maximization counts by participant. For this our workhorse will be the Mann-Whitney-Wilcoxon (MWW) test. In addition to p-values, we report an effect size for each instance of the MWW test. Given two groups with sample sizes n_1 and n_2, respectively, and a MWW test statistic value U, the effect size is given by $r = U/(n_1 n_2)$. This is the estimate of $Pr(x_2 > x_1) + \frac{1}{2}Pr(x_2 = x_1)$, where x_i is a randomly drawn individual from group $i = 1, 2$.

We look at the relationship between individual characteristics of our participants and their rates

[7] One participant declined to disclose their gender.

[8] Because our participants are drawn from across the student body and therefore a wide range of degree courses, we omit degree course here. We do look subsequently at the performance of students of economics, of whom there are few and are approximately equally scattered across the treatments.

[9] A more detailed summary of distributions of choices broken down by each of the seven sets is in Appendix A.

Subsample	Group	N	All	Maximizing 0/2	1/2	2/2
HAA results	**Post**	61	95.1	0.0	9.8	90.2
HAA results	**Pre+Amt**	50	66.0	20.0	28.0	52.0
Treatment	**Post**	49	79.6	10.2	20.4	69.4
Treatment	**Pre+Amt**	54	60.2	22.2	35.2	42.6
Treatment	**Post+Amt**	49	73.5	10.2	32.7	57.1
Treatment	**Pre+Amt+Instr**	54	70.4	18.5	22.2	59.3
Prior experience	0 to 5 sessions	112	70.1	16.1	27.7	56.3
Prior experience	10 or more sessions	94	71.3	14.9	27.7	57.5
Math-confident	No	86	59.3	22.1	37.2	40.7
Math-confident	Yes	106	78.8	12.2	17.9	69.8
Math-confident	Decline to say	14	78.6	0.0	42.9	57.1
Gender	Female	114	66.2	19.3	29.0	51.8
Gender	Male	91	76.4	11.0	25.3	63.8
Native English speaker	No	73	66.4	19.2	28.8	52.1
Native English speaker	Yes	130	74.2	12.3	26.9	60.8
Degree course	Not economics	179	67.6	16.8	31.3	52.0
Degree course	Economics	25	90.0	8.0	4.0	88.0

Table 2: Low-spade maximization rates, disaggregated by treatments and demographic characteristics. The first two rows report the comparable results from HAA. Not all demographics breakouts of our data add up to $N = 206$ due to blank responses.

Treatment	Gender = Female					Gender = Male						
			Maximizing					Maximizing				
	N	All	0/2	1/2	2/2	N	All	0/2	1/2	2/2	r	p
Post	24	70.8	16.7	25.0	58.3	22	88.6	4.6	13.6	81.8	.377	.077
Pre+Amt	21	42.9	38.1	38.1	23.8	26	69.2	15.3	30.8	53.9	.321	.025
Post+Amt	26	73.1	11.5	30.8	57.7	20	75.0	10.0	30.0	60.0	.487	.860
Pre+Amt+Instr	33	68.1	18.2	27.3	54.6	18	80.6	11.1	16.7	72.2	.412	.235
Fisher combined probability test of equality of distribution for all treatments												.047

(a) By gender

Treatment	Native English speaker = No					Native English speaker = Yes						
			Maximizing					Maximizing				
	N	All	0/2	1/2	2/2	N	All	0/2	1/2	2/2	r	p
Post	16	81.3	6.3	25.0	68.8	30	78.3	13.3	16.7	70.0	.495	.943
Pre+Amt	16	62.5	25.0	25.0	50.0	31	54.8	25.8	38.7	35.5	.444	.502
Post+Amt	13	53.8	23.1	46.2	30.8	33	81.8	6.1	24.2	69.7	.709	.013
Pre+Amt+Instr	21	61.9	28.6	19.1	52.3	30	80.0	12.9	26.6	60.5	.603	.153
Fisher combined probability test of equality of distribution for all treatments												.083

(b) By native English speaker

Table 3: Low-spade maximization rates by treatment, for characteristics without prior hypotheses. All reports overall maximization rate; $k/2$ the proportion of participants who maximized on k of the low spades trials, $k = 0, 1, 2$. r and p report the effect size and p-value, respectively, of the MWW test for equality of distribution across the characteristic, for the given treatment. The Fisher combined probability method is used to report a p-value for the joint null hypothesis of no effect due to characteristics across all treatments.

of maximization on low-spades trials. We break out the low-spades maximization data by characteristic and by treatment in Table 3 for characteristics on which we did not make a priori hypotheses (gender and native language), and in Table 4 for those on which we did make hypotheses (experience and mathematics self-confidence). Each characteristic divides our sample into two groups.[10] To look for evidence of an overall effect on maximization rates due to the characteristic, for each treatment we compare the distribution of low-spades maximization rates between the groups using MWW. Because each treatment is an independent sample, the four tests by treatment are independent tests. We aggregate the p-values of the individual tests for the four treatments using the combined probability method of Fisher (1925).

Overall, unconditional on treatment, males (76.4%) maximized more often than females (66.2%). Table 3a shows that males have a higher maximization rate in all treatments. The overall test for a gender effect is significant at the 5% level ($p = .047$).[11] Participants who are native speakers of English (74.2%) maximized more often than those who are not (66.4%). Table 3b shows a less uniform pattern across treatments. Non-native speakers are much less likely to maximize than native speakers in **Post+Amt** and **Pre+Amt+Instr**. The overall test for an effect due to native English speaking is significant at the 10% level ($p = .083$).

We observe (with admitted pleasure) that students on economics courses choose the earnings maximizing card more often (90.0%) than those on other courses (67.6%).[12] Among economics students, 18 (72.0%) identified as being good at mathematics, and 3 (12.0%) declined to say, while among students from other schools, 86 (48.0%) said they were good at mathematics, and 11 (6.2%) declined to say. These distributions are significantly different ($p = .009$ using Fisher's exact test for independence). A variety of approaches have been used in the literature as to whether to include, exclude, or control for the presence of economics students in experimental samples, often depending on the research question to hand. Friesen and Earl (2015) found that students with economics training perform better in tasks involving choices with multipart tariffs. In the UK university system, students do not in general take courses on other subjects, and so we can say that non-economics students in our sample would be unlikely to have much undergraduate-level training in economic reasoning. Nevertheless, we cannot distinguish whether the effect would be due to selection of students into the economics program, or whether training in the type of economic models that motivate experiments in economics makes students more likely to recognize and choose the earnings-maximizing card. As the economics students are small in number and spread uniformly across the four treatments, we omit this characteristic in further analysis.

We now look at the characteristics on which we set hypotheses. Participants with 10 or more

[10] In what follows, we omit the participants who did not respond to one or more of the demographics questions.

[11] We note that all sessions were led by the same female experimenter.

[12] We omit the breakout table due to the low numbers of economics students. Carrying out the same procedure as above results in a p-value of .103 for the combined probability test of no effect due to course of study.

Treatment	Prior experience $\in [0,5]$					Prior experience ≥ 10						
			Maximizing					Maximizing				
	N	All	0/2	1/2	2/2	N	All	0/2	1/2	2/2	r	p
Post	24	79.2	16.7	8.3	75.0	22	79.5	4.6	31.9	63.6	.468	.644
Pre+Amt	23	56.5	23.1	34.8	39.1	24	58.3	25.0	33.3	41.7	.513	.874
Post+Amt	27	70.4	14.8	29.6	55.6	19	78.9	5.3	31.6	63.1	.554	.484
Pre+Amt+Instr	29	74.1	10.3	31.0	58.6	22	70.5	22.7	13.6	63.6	.503	.965
Fisher combined probability test of equality of distribution for all treatments												.953

(a) By prior experience

Treatment	Math-confident = No					Math-confident = Yes						
			Maximizing					Maximizing				
	N	All	0/2	1/2	2/2	N	All	0/2	1/2	2/2	r	p
Post	18	69.4	11.1	38.9	50.0	28	85.7	10.7	7.1	82.1	.356	.044
Pre+Amt	17	55.9	23.5	41.2	35.3	30	58.3	26.7	30.0	43.3	.479	.804
Post+Amt	18	63.9	16.7	38.9	44.4	28	80.4	7.1	25.0	67.9	.376	.109
Pre+Amt+Instr	31	56.5	25.8	35.5	38.7	20	97.5	0.0	5.0	95.0	.212	.0001
Fisher combined probability test of equality of distribution for all treatments												.0002

(b) By math-confidence

Table 4: Low-spade maximization rates by treatment, for characteristics with prior hypotheses. All reports overall maximization rate; $k/2$ the proportion of participants who maximized on k of the low spades trials, $k = 0, 1, 2$. r and p report the effect size and p-value, respectively, of the MWW test for equality of distribution across the characteristic, for the given treatment. The Fisher combined probability method is used to report a p-value for the joint null hypothesis of no effect due to characteristics across all treatments.

sessions of experience maximized on 71.3% of trials overall, compared to 70.1% among those with 5 or fewer sessions. Table 4a shows that the two groups have similar maximization rates for each treatment; the overall combined probability test returns a p-value of .953, indicating no evidence of any systematic effect.

The maximization rate for those reporting they were good at math is 78.8%, compared to 59.3% for those saying they were not. A total of 14 participants declined to answer this question. Their overall maximization rate is 78.6%, suggesting they are more similar to those answering the question yes than those answering no; we drop these 14 from the analyses below. Table 4b shows that the math-confident participants maximize much more often in three of the four treatments, with the notable exception being **Pre+Amt**. The overall combined probability test results in a p-value of .0002, indicating strong evidence that the answer to this question predicts low-spade maximization behavior.

We begin the statement of the formal results following from our hypothesis by evaluating our replication of HAA's results in our overall participant pool.

Result 1 *We replicate HAA's result that the maximization rate on low-spades trials is lower in **Pre+Amt** than in **Post**. The maximization rate in **Pre+Amt** is comparable in our data and HAA, but we observe a significantly lower maximization rate in **Post**. The magnitude of the treatment effect is therefore smaller in our data overall.*

Support. As we do not have demographics breakouts for HAA's participants, we test Hypothesis 1 using aggregate data for all participants irrespective of demographics. Table 2 reports an overall low-spades maximization rate of 79.6% in **Post**, which decreases to 60.2% in **Pre+Amt**. We reject the null hypothesis that the distributions of maximization rates are the same in both treatments (MWW, $r = .639$, $p = .0069$).

The overall maximization rates on low-spades trials in **Pre+Amt** are similar in our data (60.2%) and HAA (66.0%); the distributions are not statistically different (MWW, $r = .543$, $p = .41$). Our results differ significantly in **Post**, where HAA report a 95.1% maximization rate as opposed to our 79.6% (MWW, $r = .609$, $p = .0042$). □

As shown above, there are patterns in maximization rates when participants are broken out in groups by individual characteristics. We can therefore do a more detailed analysis of our data for **Post** and **Pre+Amt**, as summarized in Table 5a. For each group we test the null hypothesis that the maximization rate is the same in **Post** and **Pre+Amt**, which generates the effect size r and p-value reported in the Table. The two tests on the groups generated by a characteristic are independent; for each characteristic we aggregate the results of the two tests using Fisher's method. Using this approach, we find the evidence for Result 1 is robust to breaking out the data by each of the four characteristics.

Result 2 *The mere mention of the total value of spades does not significantly affect behavior in postpayment conditions. The largest effect size is observed among non-native speakers of English, who maximize less often when the total value of spades is mentioned than when it is not.*

Support. Overall we observe a low-spade maximization rate of 73.5% in treatment **Post+Amt** as opposed to the 79.6% rate observed in **Post**. These overall rates are not statistically different (MWW, $r = .555$, $p = .27$). Table 5b reports on the results of this comparison applied to each group generated by individual characteristics, in which we see that no treatment difference is established when looking across any of the characteristics. The largest effect size, and smallest p-value, is found among non-native speakers of English ($r = .704$, $p = .039$); this group's maximization rate drops from 81.3% in **Post** to 53.8% in **Post+Amt**, whereas native English speakers do not show a similar pattern. This p-value is not small enough to obtain significance at the 10% level, using the Holm-Bonferroni method to account for the multiple tests. □

Result 3 *The alternate instructions rephrasing the link between each set and payment lead to higher maximization rates, comparable to the baseline postpayment treatment.*

Support. Overall, the low-spade maximization rate in **Pre+Amt+Instr** is 70.4%, which is between the 79.6% observed in **Post** and 60.2% in **Pre+Amt**. However, in Table 4b we note that in **Pre+Amt+Instr** a majority of participants did not consider themselves good at math, which is the opposite pattern from the other three treatments. The results of tests controlling for each individual characteristic are shown in Table 5c for **Pre+Amt** versus **Pre+Amt+Instr**, and Table 5d for **Post** versus **Pre+Amt+Instr**. The null hypothesis of no difference in distribution is not rejected for any of the tests between **Post** and **Pre+Amt+Instr**. On the other hand, the null hypothesis is strongly rejected when controlling for math-confidence when comparing **Pre+Amt** and **Pre+Amt+Instr** ($p = .0020$), as well as rejected at 5% when controlling for gender ($p = .038$) and 10% when controlling for native English speaking ($p = .051$). □

Our formulation of Hypothesis 2 specifically related to the mention of spades under the postpayment condition only, and Hypothesis 3 the effects of rephrasing instructions under prepayment. All treatments other than **Post** involved a mention of the total value of spades in the instructions. Ex post, we observe in Table 3b that the maximization rate for non-native speakers is more than 20 percentage points lower across these three treatments in which the total value of spades is mentioned, compared to **Post**. While **Pre+Amt+Instr** results in an increase in maximization for native English speakers compared to **Pre+Amt**, a similar effect is not evident among non-native speakers.

Characteristic	Group	r	p	Group	r	p	Combined
Gender	Male	.643	.041	Female	.688	.021	.0069
Native English speaker	Yes	.668	.013	No	.617	.198	.018
Math-confident	Yes	.688	.0046	No	.596	.292	.011
Experience	$\geq$ 10 sessions	.642	.069	0-5 sessions	.661	.033	.016

(a) **Post** versus **Pre+Amt**

Characteristic	Group	r	p	Group	r	p	Combined
Gender	Male	.609	.127	Female	.492	.913	.365
Native English speaker	Yes	.490	.871	No	.704	.039	.150
Math-confident	Yes	.561	.304	No	.539	.665	.525
Experience	$\geq$ 10 sessions	.504	.963	0-5 sessions	.579	.256	.592

(b) **Post** versus **Post+Amt**

Characteristic	Group	r	p	Group	r	p	Combined
Gender	Male	.412	.259	Female	.329	.025	.038
Native English speaker	Yes	.323	.0090	No	.500	1.000	.051
Math-confident	Yes	.235	.0002	No	.494	.945	.0020
Experience	$\geq$ 10 sessions	.411	.257	0-5 sessions	.380	.107	.126

(c) **Pre+Amt** versus **Pre+Amt+Instr**

Characteristic	Group	r	p	Group	r	p	Combined
Gender	Male	.552	.444	Female	.519	.786	.717
Native English speaker	Yes	.504	.942	No	.612	.191	.489
Math-confident	Yes	.433	.172	No	.587	.279	.194
Experience	$\geq$ 10 sessions	.533	.660	0-5 sessions	.560	.370	.589

(d) **Post** versus **Pre+Amt+Instr**

Table 5: Summary of tests comparing low-spade maximization rates across treatments. MWW is used for the individual test of each group. Effect sizes r are reported such that $r > .5$ corresponds to a higher rate in the treatment listed first. Combined gives the p-value of obtained by aggregating the results of the MWW tests for the corresponding groups using Fisher's combined probability test.

We now look more specifically at the two individual characteristics on which we had set prior hypotheses.

Result 4 *We do not find any evidence that participants who have participated in 10 or more previous sessions are more or less likely to maximize earnings than those who have been in 1 to 5 sessions previously.*

Support. Table 4a presents MWW tests of the null hypothesis of no difference in maximizing rates between the two experience groups. The null hypothesis is not rejected for any treatment; p-values are uniformly high and effect sizes are uniformly close to .5, corresponding to no effect. Of the four individual characteristics considered, breakouts by experience show the least between-group variation within each treatment, as well as the least evidence for any differential treatment effects. □

Result 5 *Reported mathematics confidence overall is a strong predictor of the likelihood to choose earnings-maximizing cards in low-spades trials. However, the treatment difference in **Post** versus **Pre+Amt+Instr** is accounted for mainly by its effect among participants who identify as being good at math.*

Support. Table 4b presents MWW tests of the null hypothesis of no difference in maximizing rates. Overall, there is evidence of a strong effect correlating to the answer to this question ($p = .0002$). The effect is not uniform across treatments. In **Post**, those identifying as good at math maximize significantly more often, by an 85.7% to 69.4% margin; the MWW test shows this difference is significant ($r = .356$, $p = .044$). Likewise there is a difference between the groups in **Pre+Amt+Instr**, with math-confident participants maximizing on 97.5% of trials and others on 56.5% (MWW $r = .212$, $p = .0001$). However, in **Pre+Amt** the performance of the two groups is not distinguished; math-confident participants maximize on 58.3% of trials and others on 55.9% (MWW $r = .479$, $p = .804$). □

The positive correlation overall between math-confidence and performance on the task is intuitive, although it is interesting that it comes through so strongly on a task which might not seem that demanding at first look. More surprising is the result that it is math-confident participants whose maximization rates lower most sharply in **Pre+Amt**. When we look only at math-confident participants, the size of our treatment effect between **Post** and **Pre+Amt** is comparable to the results reported by HAA. We observe a drop in maximization rate from 85.7% to 58.3% in this group, and HAA report a drop from 95.1% to 66.0%. HAA conducted their experiments at Duke University, a private institution that is one of the most selective in the United States. Our replication was done at a middle-tier public university in the United Kingdom, which is best known for its degree programs in areas such as literature and creative writing, and which offers no engineering or physics courses at all. It is plausible that our participants who identify as math-confident are more comparable to the sample at Duke; if that is the case, our results provide a firm replication of HAA's effect.

Our results seem rule out an account in which the treatment effect between **Post** and **Pre+Amt** occurs due to less-numerate participants making more frequent calculation errors in **Pre+Amt**. Such a hypothesis would be supported by a larger drop in the maximization rate of participants with less math-confidence; we find exactly the opposite. The lack of a large treatment effect among less math-confident participants may arise because of the extra noise due to a higher error rate in the arithmetic calculations required; the higher baseline maximization rate of math-confident participants in **Post** makes it easier to pick up a treatment effect.

Although math-confident participants may not be making arithmetic errors, they may still be making errors in judgment in **Pre+Amt** due to other factors.[13] If this is the case, then alternative phrasings like those in **Pre+Amt+Instr** may have the effect of debiasing the decision-maker. These effects have been explored in psychology (e.g. Evans et al., 1994), law (e.g. Babcock et al., 1997), and accounting (e.g. Clarkson et al., 2002) as well as economics (e.g. List, 2001). In the case of the instructions used in Treatment **Pre+Amt+Instr**, explicitly stepping through the process by which payments will be realized might influence whatever processes participants use to come to their decisions.

4 Conclusion

HAA reported a striking result: in a simple choice task with no objective risk, participants were less likely to choose earnings-maximizing cards when they received a prepayment based on some other, non-earnings-maximizing choice. We have shown this effect is qualitatively robust, even with a different participant pool and somewhat higher incentives.

Given the apparently straightforward calculations required to maximize earnings in this decision task, it is remarkable that undergraduate students both at Duke (HAA) and University of East Anglia (this paper) would, overall, leave money on the table. We therefore take a closer look at this decision task, both to map out the extent to which maximization rates depend on details of the experimental protocol, and to understand better whether certain subgroups of participants are less likely to maximize earnings.

We find that the effect of prepayment does depend on the description of the experimental task. The methodological observation that participant decisions are not independent of instructions is

[13]For the purposes of this paper, we put to one side the question of normative implications. If a decision-maker is truly loss-averse, then choosing the non-earnings-maximizing card might be viewed as optimal from the perspective of their preferences. McQuillin and Sugden (2012) provide one survey of the issues in reconciling behavioral and normative economics.

hardly novel, but the potential practical implications are noteworthy. We can think of the description of the experimental task as the analog of the description of offers to consumers or terms and conditions of agency contracts. One interpretation of HAA's results and our replication is that it is easy to influence decisions, even when there is no objective risk and the choice architecture is not very complex. HAA's instructions, used in the **Pre+Amt** treatment, while brief, are not overtly misleading. Our **Pre+Amt+Instr** instructions are somewhat more explicit, which may serve a debiasing function. A regulator whose objective was to keep people from making apparently suboptimal decisions might find descriptions similar to ours preferable because of their debiasing effects. However, creating effective written communication is challenging.[14] Because instructions like HAA's are not overtly misleading, and because the effect of written communication on an audience is difficult to judge, even by experienced writers, the regulator might find it difficult to distinguish between well-intentioned but unclear writing, as opposed to deliberate obfuscation.

This simple choice task devised by HAA generates an interesting amount of variation in performance, depending on how it is presented, and on the characteristics of the decision-makers. The task is sufficiently stylized that broad claims about external validity should be tempered. However, the stark simplicity of the setting helps to illustrate some possible subtle mechanisms firms could use to influence the behavior of agents or consumers, with corresponding implications for the problems that market designers or regulators face in evaluating the appropriateness of the structure of contracts or offers.

References

J. Abeler and D. Nosenzo. Self-selection into laboratory experiments: Pro-social motives versus monetary incentives. *Experimental Economics*, 18:195–214, 2014.

H. Allcott. Consumers' perceptions and misperceptions of energy costs. *American Economic Review*, 101: 98–104, 2011.

M. H. Ashcraft. Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, 11:181–185, 2002.

M. H. Ashcraft and E. P. Kirk. The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology – General*, 130:224–237, 2001.

M. H. Ashcraft and K. S. Ridley. Math anxiety and its cognitive consequences. In J. I. D. Campbell, editor, *Handbook of Mathematical Cognition*, pages 315–327. Psychology Press, New York, 2005.

L. Babcock, G. Loewenstein, and S. Issacharoff. Creating convergence: Debiasing biased litigants. *Law & Social Inquiry*, 22:913–925, 1997.

[14] Pinker (1994, 2014) discusses some reasons why writing for clear communication is difficult, and even writers sincerely attempting to be clear may not be successful.

A. Bandura. *Social Learning Theory*. General Learning Press, New York, 1977.

B. M. Barber and T. Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 116, 2001.

S. W. Becker, J. Ronen, and G. H. Sorter. Opportunity costs - an experimental approach. *Journal of Accounting Research*, 12:317–329, 1974.

O. Bock, A. Nicklisch, and I. Baetge. hRoot: Hamburg registration and organization online tool. WiSo-HH Working Paper Series, Number 1, 2012.

T. Buser, M. Niederle, and H. Oosterbeek. Gender, competitiveness, and career choices. *Quarterly Journal of Economics*, 129:1409–1447, 2014.

P. M. Clarkson, C. Emby, and V. W. S. Watt. Debiasing the outcome effect: The role of instructions in an audit litigation setting. *Auditing: A Journal of Practice & Theory*, 21:7–20, 2002.

A. Conte, M. V. Levati, and N. Montinari. Experience in public goods experiments. Jena Economic Research Papers, Number 2014-010, 2014.

J. Eccles. Perceived control and the development of academic motivation: Commentary. *Monographs of the Society for Research in Child Development*, 63:221–231, 1998.

J. S. T. Evans, S. E. Newstead, J. L. Allen, and P. Pollard. Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology*, 6:263–285, 1994.

U. Fischbacher. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10:171–178, 2007.

Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925.

L. Friesen and P. E. Earl. Multipart tariffs and bounded rationality: An experimental analysis of mobile phone plan choices. *Journal of Economic Behavior & Organization*, 116:239–253, 2015.

A. Gaudeul and R. Sugden. Spurious complexity and common standards in markets for consumer goods. *Economica*, 79:209–225, 2012.

M. D. Grubb. Failing to choose the best price: Theory, evidence, and policy. *Review of Industrial Organization*, 47:303–340, 2015.

M. D. Grubb and M. Osborne. Cellular service demand: Biased beliefs, learning, and bill shock. *American Economic Review*, 105:234–271, 2015.

G. Hochman, S. Ayal, and D. Ariely. Keeping your gains close but your money closer: The prepayment effect in riskless choices. *Journal of Economic Behavior & Organization*, 107B:582–594, 2014. doi: 10.1016/j.jebo.2014.01.014.

D. Kahneman, J. L. Knetsch, and R. H. Thaler. Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, pages 1325–1348, 1990.

D. Kahneman, J. L. Knetch, and R. H. Thaler. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5:193–206, 1991. doi: 10.1257/jep.5.1.193.

K. Kalayci and J. Potters. Buyer confusion and market prices. *International Journal of Industrial Organization*, 29:14–22, 2011.

R. P. Larrick and J. B. Soll. The MPG illusion. *Science*, 3230:1593–1594, 2008.

J. List. Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. *American Economic Review*, 91:1498–1507, 2001.

P. Lunn. Telecommunications consumers: A behavioral economic analysis. ESRI Working Paper 417, 2011.

H. Marsh, K. Xu, and A. J. Martin. Self-concept: A synergy of theory, method, and application. In K. Harris, S. Graham, and T. Urdan, editors, *APA Educational Psychology Handbook, Volume 1: Theories, Constructs, and Critical Issues*, pages 427–548. American Psychological Association, Washington DC, 2012.

A. Matthey and T. Regner. On the independence of history: Experience spill-overs between experiments. *Theory and Decision*, 75:403–419, 2013.

B. McQuillin and R. Sugden. Reconciling normative and behavioural economics: The problems to be solved. *Social Choice and Welfare*, 38:553–567, 2012.

M. Niederle and L. Vesterlund. Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122:1067–1101, 2007.

F. Pajares and M. D. Miller. Role of self-efficiacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86:193–203, 1994.

M. Piccione and R. Spiegler. Price competition under limited comparability. *Quarterly Journal of Economics*, 127:97–135, 2012.

S. Pinker. *The Language Instinct*. Harper Perennial Modern Classics, New York, 1994.

S. Pinker. *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century*. Penguin, New York, 2014.

A. Schram. Artificality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12:225–237, 2005.

S. Sitzia and D. J. Zizzo. Does product complexity matter for competition in experimental retail markets. *Theory and Decision*, 70:65–82, 2011.

S. Sitzia, J. Zheng, and D. J. Zizzo. Inattentive consumers in markets for services. *Theory and Decision*, 79:307–322, 2015.

R. H. Thaler. Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1:39–60, 1980.

R. H. Thaler. Mental accounting matters. *Journal of Behavioral Decision Making*, 12:183–206, 1999.

C. M. Wilson and C. Waddams-Price. Do consumers switch to the best supplier? *Oxford Economic Papers*, 62:647–668, 2010.

A Choice distributions at the set level

In this appendix we provide a more detailed view on the seven sets of cards faced by the participants. In each table, the first row shows the four cards that were used in the set. The next row contains the values of the cards; the data for the card(s) which maximize earnings are displayed in bold. The total number of choices and corresponding frequencies for each card are presented for each treatment, **Post**, **Post+Amt**, **Pre+Amt**, and **Pre+Amt+Instr**, in separate rows, recalling that **Post** and **Pre+Amt** are replications of HAA.

Set 1	Q♡		4♣		6♢		6♡	
Value	**GBP 1.20**		GBP 0.40		GBP 0.60		GBP 0.60	
Post	**94%**	**46**	4%	2	2%	1	0%	0
Post+Amt	**88%**	**43**	8%	4	2%	1	2%	1
Pre+Amt	**96%**	**52**	4%	2	0%	0	0%	0
Pre+Amt+Instr	**87%**	**47**	6%	3	7%	4	0%	0

Set 2	2♡		3♢		5♠		3♡	
Value	GBP 0.20		GBP 0.30		**GBP 1.25**		GBP 0.30	
Post	0%	0	0%	0	**100%**	**49**	0%	0
Post+Amt	0%	0	6%	3	**94%**	**46**	0%	0
Pre+Amt	0%	0	2%	1	**98%**	**53**	0%	0
Pre+Amt+Instr	0%	0	7%	4	**93%**	**50**	0%	0

Set 3	J♢		4♠		7♡		9♡	
Value	**GBP 1.10**		GBP 1.00		GBP 0.70		GBP 0.90	
Post	**73%**	**36**	27%	13	0%	0	0%	0
Post+Amt	**71%**	**35**	29%	14	0%	0	0%	0
Pre+Amt	**54%**	**29**	46%	25	0%	0	0%	0
Pre+Amt+Instr	**65%**	**35**	30%	16	2%	1	4%	2

Set 4	J♣		K♢		7♢		6♣	
Value	GBP 1.10		**GBP 1.30**		GBP 0.70		GBP 0.60	
Post	2%	1	**98%**	**48**	0%	0	0%	0
Post+Amt	6%	3	**92%**	**45**	0%	0	2%	1
Pre+Amt	2%	1	**96%**	**52**	0%	0	2%	1
Pre+Amt+Instr	2%	1	**96%**	**52**	2%	1	0%	0

Set 5	4♡		10♢		A♠		8♡	
Value	GBP 0.40		**GBP 1.00**		GBP 0.25		GBP 0.80	
Post	0%	0	**86%**	**42**	14%	7	0%	0
Post+Amt	0%	0	**76%**	**37**	22%	11	2%	1
Pre+Amt	0%	0	**67%**	**36**	31%	17	2%	1
Pre+Amt+Instr	0%	0	**76%**	**41**	20%	11	4%	2

Set 6	3♠		2♣		A♢		4♢	
Value	**GBP 0.75**		GBP 0.20		GBP 0.10		GBP 0.40	
Post	**86%**	**42**	0%	0	4%	2	10%	5
Post+Amt	**67%**	**33**	0%	0	12%	6	20%	10
Pre+Amt	**87%**	**47**	2%	1	0%	0	11%	6
Pre+Amt+Instr	**80%**	**43**	2%	1	6%	3	13%	7

Set 7	5♢		2♢		3♣		2♠	
Value	**GBP 0.50**		GBP 0.20		GBP 0.30		**GBP 0.50**	
Post	**53%**	**26**	0%	0	2%	1	**45%**	**22**
Post+Amt	**41%**	**20**	0%	0	0%	0	**59%**	**29**
Pre+Amt	**7%**	**4**	0%	0	0%	0	**93%**	**50**
Pre+Amt+Instr	**15%**	**8**	0%	0	0%	0	**85%**	**46**

Online supplementary material for
Mathematics self-confidence and the "prepayment effect" in riskless choices
Lian Xue, Stefania Sitzia, Theodore L. Turocy

1 Comparison of instructions

In this appendix, we provide the full text of the instructions for each of the four treatments. In addition, we provide the instructions as used by Hochman et al. (2014) for comparison.

1.1 Treatment Post

1.1.1 This paper

This is an experiment in the economics of decision making. If you follow the instructions and make appropriate decisions, you can earn an appreciable amount of money. You will receive your earnings for today's session in cash before you leave the laboratory.

It is important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

We will now describe the session in more detail. Please follow along with these instructions as they are read aloud.

On your computer monitor, you will be given sets of four playing cards, face-up, several times. Each time, you will select and keep one card.

After all of these sets are completed, you will be paid based on the cards you have selected. Each card is worth its point value times 10p, so a three is worth three times 10p, a nine is worth nine times 10p, et cetera. Aces are worth 10p, jacks are worth eleven times 10p, queens are worth twelve times 10p, and kings are worth thirteen times 10p.

However, those values apply only to cards that are NOT spades. Spades (♠) are worth their point value times 25p, not 10p. The ace of spades is worth 25p, the two of spades is worth two times 25p, and so on.

At the end of the game, we will pay you for the cards you have selected.

1.1.2 HAA

Thank you for your participation. Feel free to ask questions at any time if anything is unclear. There are no tricks or catches to this game, we simply ask that you pay attention to the instructions and think carefully about your decisions. You will be paid some amount of money at the end of the game; how much you are paid will be determined by the decisions you make.

You will be given sets of four playing cards, face-up, several times. Each time, you will select and keep one card.

After all of these sets are completed, you will be paid based on the cards you have selected.

Each card is worth its point value in dimes, so a three is worth three dimes, a nine is worth nine dimes, et cetera. Aces are worth one dime, jacks are worth eleven dimes, queens are worth twelve, and kings are worth thirteen.

However, those values apply only to cards that are NOT spades. Spades are worth their point value in quarters, not dimes. The ace of spades is worth one quarter, the two of spades is worth two quarters, and so on. We have placed the cards ace through five of spades in the deck randomly.

At the end of the game, we will pay you for the cards you have selected.

1.2 Treatment Post+Amt

This is an experiment in the economics of decision making. If you follow the instructions and make appropriate decisions, you can earn an appreciable amount of money. You will receive your earnings for today's session in cash before you leave the laboratory.

It is important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

We will now describe the session in more detail. Please follow along with these instructions as they are read aloud.

On your computer monitor, you will be given sets of four playing cards, face-up, several times. Each time, you will select and keep one card.

After all of these sets are completed, you will be paid based on the cards you have selected.

Each card is worth its point value times 10p, so a three is worth three times 10p, a nine is worth nine times 10p, et cetera. Aces are worth 10p, jacks are worth eleven times 10p, queens are worth twelve times 10p, and kings are worth thirteen times 10p.

However, those values apply only to cards that are NOT spades. Spades (♠) are worth their point value times 25p, not 10p. The ace of spades is worth 25p, the two of spades is worth two times 25p, and so on. We have placed the cards ace through five of spades in the deck randomly. The value of the five spades, ace through five, equals a total of three pounds and seventy-five pence: one plus two plus three plus four plus five is fifteen times 25p i.e. three pounds and seventy-five pence.

At the end of the game, we will pay you for the cards you have selected.

1.3 Treatment Pre+Amt

1.3.1 This paper

This is an experiment in the economics of decision making. If you follow the instructions and make appropriate decisions, you can earn an appreciable amount of money. You will receive your earnings for today's session in cash before you leave the laboratory.

It is important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

We will now describe the session in more detail. Please follow along with these instructions as they are read aloud.

On your computer monitor, you will be given sets of four playing cards, face-up, several times. Each time, you will select and keep one card.

After all of these sets are completed, you will be paid based on the cards you have selected. Each card is worth its point value times 10p, so a three is worth three times 10p, a nine is worth nine times 10p, et cetera. Aces are worth 10p, jacks are worth eleven times 10p, queens are worth twelve times 10p, and kings are worth thirteen times 10p.

However, those values apply only to cards that are NOT spades. Spades (♠) are worth their point value times 25p, not 10p. The ace of spades is worth 25p, the two of spades is worth two times 25p, and so on. We have placed the cards ace through five of spades in the deck randomly.

The value of the five spades (ace through five) equals a total of three pounds and seventy-five pence: one plus two plus three plus four plus five is fifteen times 25p i.e. three pounds and seventy-five pence. We will give you this amount up front.

However, if you do not choose all of the five spade cards, you will need to give us back some of this money at the end of the game. The amount you return will be the value of the spade card(s) that you did NOT choose.

For example, if you do not pick up the three of spades, you will return three times 25p to us from your three pounds and seventy-five pence.

At the end of the game, if you have not selected all spades, we will pay you for the cards you have selected, and you will refund us money for the spades you have not selected.

1.3.2 HAA

Thank you for your participation. Feel free to ask questions at any time if anything is unclear. There are no tricks or catches to this game, we simply ask that you pay attention to the instructions and think carefully about your decisions. You will be paid some amount of money at the end of the game; how much you are paid will be determined by the decisions you make.

You will be given sets of four playing cards, face-up, several times. Each time, you will select and keep one card.

After all of these sets are completed, you will be paid based on the cards you have selected. Each card is worth its point value in dimes, so a three is worth three dimes, a nine is worth nine dimes, et cetera. Aces are worth one dime, jacks are worth eleven dimes, queens are worth twelve, and kings are worth thirteen.

However, those values apply only to cards that are NOT spades. Spades are worth their point value in quarters, not dimes. The ace of spades is worth one quarter, the two of spades is worth two quarters, and so on. We have placed the cards ace through five of spades in the deck randomly.

The value of the five spades (ace through five) equals a total of three dollars and seventy-five cents: one plus two plus three plus four plus five is fifteen quarters i.e. three dollars and seventy five cents. We will give you these fifteen quarters up front.

However, if you do not choose all of the five spade cards, you will need to give us back some of this money at the end of the game. The amount you return will be the value of the spade card(s) that you did NOT choose.

For example, if you do not pick up the three of spades, you will return three quarters to us from your three dollars and seventy-five cents.

At the end of the game, if you have not select all spades, we will pay you for the cards you have selected, and you will refund us money for the spades you have not selected.

1.4 Treatment Pre+Amt+Instr

This is an experiment in the economics of decision-making. If you follow the instructions and make appropriate decisions, you can earn an appreciable amount of money. You will receive your earnings for today's session in cash before you leave the laboratory.

It is important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

We will now describe the session in more detail. Please follow along with these instructions as they are read aloud.

On your computer monitor, you will be given sets of four playing cards, face-up, several times. No card will appear more than one time during the experiment. Each time, you will select and keep one card. After all of these sets are completed, you will be paid based on the cards you have selected.

The value of a card to you depends on the card's suit and point value. Spades (♠) are worth their point value times 25p. The ace of spades is worth 25p, the two of spades is worth two times 25p, and so on.

We have placed the cards ace through five of spades in the deck randomly. The values of these five spades (ace through five) equal a total of three pounds and seventy-five pence, because one plus two plus three plus four plus five is fifteen times 25p is three pounds and seventy-five pence. We will give you this amount up front.

Cards of the other suits, hearts (♡), diamonds (♢), and clubs (♣), are worth their point value times 10p. So, a three is worth three times 10p, a nine is worth nine times 10p, et cetera. In these suits, aces are worth 10p, jacks are worth eleven times 10p, queens are worth twelve times 10p, and kings are worth thirteen times 10p.

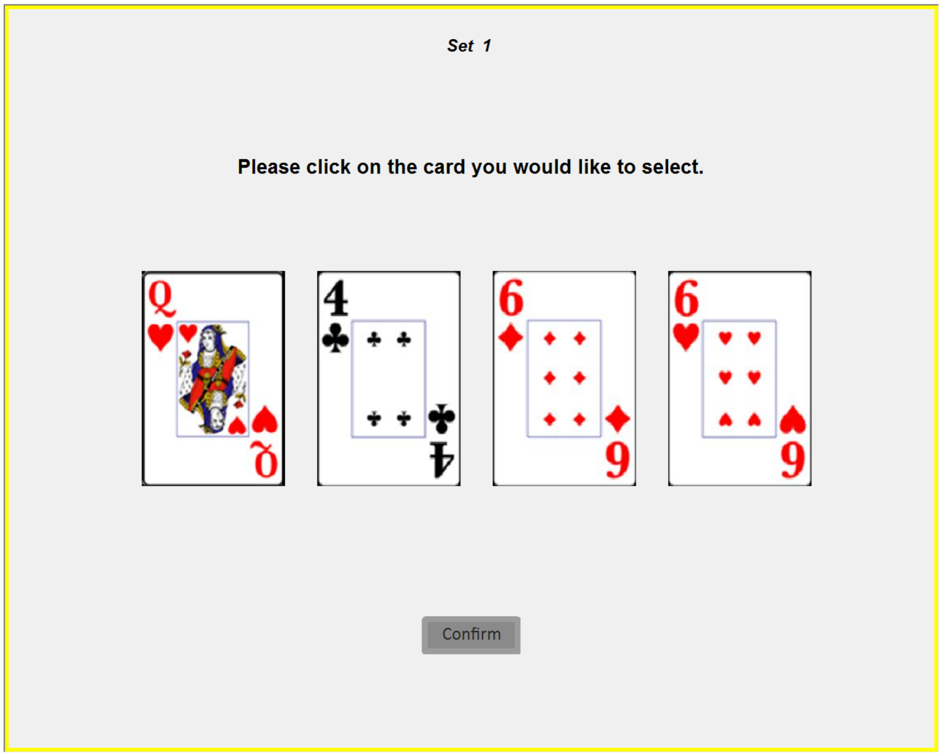
At the end of the experiment, you will leave the lab with a total payment equal to the sum of the values of the cards you select. For each set of four playing cards, there are three possible scenarios:

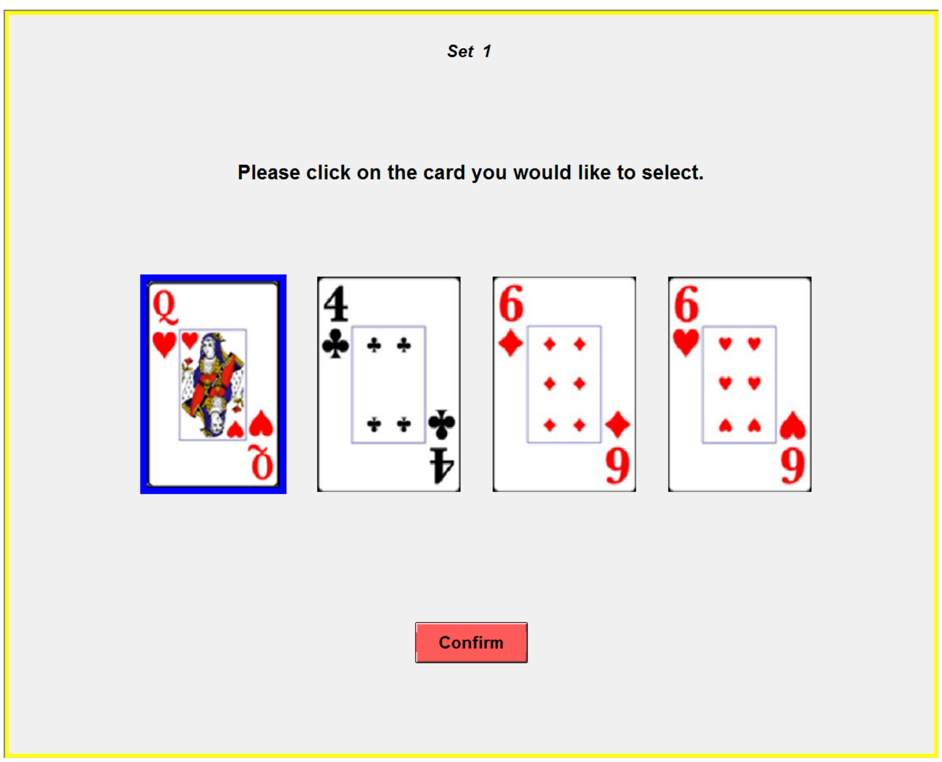
- There is not a spade among the four cards. Then, at the end of the session, you will receive a payment for that set equal to the value of the card you select.

- There is a spade among the four cards, but you select a different card. Then, at the end of the session, you will pay us back from your up front payment an amount equal to the value of the spade, and you will receive a payment from us equal to the value of the card you did select. For example, if you did not select three of spades but instead select four of diamonds; we will pay you 40p corresponding to the four of diamonds and you need to refund us 3 times 25p corresponding to the three of spades.

- There is a spade among the four cards, and you select the spade. Then, because you have already received payment for that spade card in your up front payment, you will not pay us back anything for that set, nor will you receive any additional payment for that set.

2 Screenshots of participant interface





Earnings Summary

Set	Prepaid Card	Your Selection	You need to return	We will pay you
1	—	Q♥	—	£ 1.20
2	5♠	5♠	—	—
3	4♠	J♦	£ 1.00	£ 1.10
4	—	K♦	—	£ 1.30
5	A♠	10♦	£ 0.25	£ 1.00
6	3♠	3♠	—	—
7	2♠	5♦	£ 0.50	£ 0.50
Totals	**£ 3.75**		£ 1.75	£ 5.10