Game form recognition in preference elicitation, cognitive abilities and cognitive load

Drichoutis, Andreas and Nayga, Rodolfo

Agricultural University of Athens, University of Arkansas

3 November 2019

Game form recognition in preference elicitation, cognitive abilities and cognitive load*

Andreas C. Drichoutis†[1] and Rodolfo M. Nayga, Jr.‡[2]

[1]Agricultural University of Athens
[2]University of Arkansas

First Draft: November 3, 2019
This Version: January 6, 2020

Abstract: This study further examines the failure of game form recognition in preference elicitation (Cason and Plott, 2014) by making elicitation more cognitively demanding through a cognitive load manipulation. We hypothesized that if subjects misperceive one game for another game, then by depleting their cognitive resources, subjects would misconceive the more-cognitively demanding task for the less-cognitively demanding task at a higher rate. We find no evidence that subjects suffer from a first-price-auction game-form misconception but rather that once cognitive resources are depleted, subjects' choices are better explained by random choice. More cognitively able subjects are more immune to deviations from sub-optimal play than lower cognitively able subjects.

Keywords: Game form recognition, game form misconception, Becker-DeGroot-Marschak mechanism, first price auction, preference elicitation, cognitive load, cognitive resources, Raven test.

JEL codes: C80, C91, D44

1 Introduction

Competing theories of neoclassical economics posit that preferences depend on the context faced by an individual in a choice situation. In a well cited paper, Cason and Plott (2014)

*We would like to thank Tim Cason, Achilleas Vassilopoulos for helpful suggestions; Effrosyni Papadopoulou for excellent research assistance.

†Associate Professor, Department of Agricultural Economics & Rural Development, School of Applied Economics and Social Sciences, Agricultural University of Athens, Iera Odos 75, 11855, Greece, e-mail: adrihout@aua.gr.

‡Distinguished Professor and Tyson Endowed Chair, Department of Agricultural Economics & Agribusiness, University of Arkansas, Fayetteville, AR 72701, USA, tel:+1-4795752299 e-mail: rnayga@uark.edu.

argue that because of mistakes, choices are masked as evidence of non-standard preferences. To demonstrate how decision errors (which Cason and Plott (2014) call failure of game form recognition - GFR or game form misconception - GFM) can mistakenly be regarded as subjects having non-standard preferences, acting irrationally, or being affected by framing, they performed an induced value experiment using the Becker-DeGroot-Marschak (BDM) mechanism (Becker et al., 1964).

In their experiment, Cason and Plott (2014) asked student subjects to state an offer price for selling back to the experimenter an endowed card worth \$2 to the students. If the offer price was lower or equal to a randomly drawn price, then the subject would receive the randomly drawn price; otherwise the subject would redeem the card for \$2. However, they found that only 16.7% of subjects chose offers within 5 cents of the \$2 induced value. Moreover, they showed that the mistakes were not simply random departures from a correct understanding of the experimental task, but rather they arose from a misconception of the rules of the BDM mechanism. Based on the observed data patterns, they concluded that some subjects believed that the lowest offer wins and would be paid the offer price; that is, they misconceived the BDM mechanism for a first price auction (FPA). Cason and Plott (2014) acknowledge that many other misconceptions are possible but that the FPA-GFM is consistent with what some subjects stated they should be paid.

Following up on Cason and Plott (2014), Bull et al. (2019) suggest an additional test of the game form misconception. If subjects do not bid optimally in the BDM mechanism because they misconceive it for a FPA, then they should bid optimally in the FPA. They find that subjects treated the two games as the same task when the tasks were presented simultaneously and despite the fact that a warning was given to them that the tasks were different. Thus, the simultaneous presentation of the two tasks did not result in better GFR.

Li (2017) provides a somewhat related explanation for the divergence of bids from the theoretical equilibrium value of a valuation task: although a mechanism may be strategy-proof (i.e., the weakly-dominant strategy of every bidder is to reveal their private values), the mechanism may be cognitively complex which would render it a *non-obviously* strategy-proof task. Complexity of a mechanism could vary between subjects depending on whether the subjects can easily understand (or not) the mechanics of the task. Hence, the ability of the mechanism to reveal an agent's preference could depend more on the complexity of the mechanism for low cognitive ability subjects. Related to this last point, Hassidim et al. (2017) report that many studies find that individuals misrepresent their preferences at a higher rate when they are of lower cognitive ability. In addition, Lee et al. (2017) find that subjects of higher cognitive ability tend to bid closer to their induced value in a second price auction (SPA) and large overbids are vastly the typical behavior of subjects with low cognitive ability. Li (2017) also compares the SPA and the ascending clock auction (which is strategically equivalent to the SPA) and finds that subjects play the dominant strategy at significantly higher rates under the *obviously* strategy-proof ascending clock auction, compared to the SPA which is just strategy-proof (but not *obviously* strategy-proof).[1]

It is then possible that low cognitive ability can be a detrimental factor in truthful preference revelation, consistent with the view in Choi et al. (2014) that the choices that some people make may be different from the choices they would make if they had the skills or knowledge to make better decisions. The role of intelligence or cognitive ability has been the subject of many studies (see for example Brañas Garza and Smith, 2016; Rustichini, 2015, and citations therein) and some of the stylized facts from this literature suggest that people of high cognitive ability are more risk-tolerant, more patient, and less prone to anchoring effects than those with lower cognitive ability (see Deck and Jahedi, 2015, and citations therein).

Similar to Bull et al. (2019), we use both a BDM and a FPA in order to test the base result of Cason and Plott (2014) that subjects in a BDM bid as if they participate in a FPA. However, because the two valuation tasks differ in their rules and complexity, subjects may require different cognitive resources to fully comprehend each of them. In order to exogenously manipulate cognitive resources, we employ a cognitive load manipulation by requiring subjects to memorize strings of different lengths while they are making choices. Imposing a burden on working memory has been shown to have adverse effects on performance in a variety of tasks that involve logic or reasoning (see Deck and Jahedi, 2015, and citations therein). Our design is not intended to test competing theories of framing like the endowment effect, anchoring, attraction to the maximum or expectations of trade that could explain data patterns. This would necessitate a different set of experiments which is beyond the scope of the main purpose of our paper which is to examine the specific GFM set forth as an explanation in Cason and Plott (2014).

Moreover, bids from the FPA allow us to test whether subjects in the FPA bid optimally. Our unique contribution is that by manipulating cognitive load, we are able to study the causal effect of cognitive resources on bidding behavior. Our hypothesis is that when less cognitive resources are available, subjects would be moved further away from truthful revelation of their induced values and that the BDM task would be more adversely affected than the FPA. Our

[1] A mechanism is obviously strategy-proof (OSP) if it has an equilibrium in obviously dominant strategies. A strategy S_i is obviously dominant if, for any deviating strategy S_i', the best possible outcome from S_i' is no better than the worst possible outcome from S_i. As can be seen in Table A7 in the Electronic Supplementary Material, neither the BDM mechanism nor the FPA are OSP mechanisms. For both mechanisms, the best possible outcome for $b \neq 5$ ($b \neq 6.5$ in the FPA) is better than the worst possible outcome for $b = 5$ ($b = 6.5$ in the FPA). Although both mechanisms are not OSP, there is a widespread belief that the price mechanism of the FPA (paying what you bid) is simpler and more transparent than the price mechanism of the BDM (paying a random draw). For example, Google Ads, an online advertising platform developed by Google where advertisers pay to display brief advertisements, services, video content etc., uses auctions to determine which search ads are displayed on the search results page as well as ad's rank position. Google Ads recently switched to an FPA from an SPA, citing reasons such as reducing complexity and increasing transparency for the switch (Bigler, 2019).

experimental design is a simple 2×2 between-subjects design where we vary the elicitation mechanism (BDM vs. FPA) and the level of cognitive load (high vs. low).

We find no evidence that subjects misperceive the BDM task for a FPA. While placing subjects under high cognitive load leads to poorer performance and larger deviations from induced value in the BDM task, this is not because subjects misconceive the BDM for a FPA at a higher rate but rather because the choice process of a bid becomes in effect a random choice. Moreover, the probability of submitting sub-optimal bids in the BDM task is larger for subjects with low cognitive ability, while subjects with high cognitive ability are largely unaffected. We also find that the FPA is rather immune to the high cognitive load treatment and the probability of submitting sub-optimal offers does not vary with the level of cognitive ability. This result corroborates well with the perception of the FPA as less cognitively demanding. Furthermore, results are robust to a number of robustness checks: integrating risk aversion in the analysis, accounting for attention/comprehension of instructions as well as to checks of whether subjects understand the payoff mechanism.

The next section describes the experimental design. In Section 3 we first present evidence of whether the cognitive load treatment was difficult enough and whether it had measurable effects on a set of unrelated tasks (i.e., we present some manipulation checks). We then fit structural models to allow us to distinguish between the optimal offer model under the BDM and the FPA offer model. We also examine how deviations from optimal offers differ with respect to the interacting effect of cognitive abilities and the cognitive load treatments. The results section concludes with a few robustness checks. We then discuss the importance of our findings and further recommendations for future research in the last section.

2 Experimental design

In May 2018, we recruited 269 subjects from the undergraduate population of the Agricultural University of Athens in Greece to participate in a computerized experiment at the Laboratory of Behavioral and Experimental Economics Science (LaBEES-Athens). Subjects were recruited using ORSEE (Greiner, 2015). Although subjects participated in group sessions, there was no interaction between subjects and the group sessions only served as a means to economize on resources. Subjects were randomly allocated to one of the cells of a 2×2 experimental design and each subject was only exposed to one of them (i.e, we did a between subjects design). Table 1 shows the number of subjects allocated to each of the treatments. The number of recruited subjects was dictated by sample size calculations that can be found in the Electronic Supplementary Material.

When subjects arrived at the lab, they were given a consent form to sign. They were then randomly assigned to one of the PC private booths. All instructions were computerized,

Table 1: Experimental design

	LCL	HCL	Total
FPA	67	66	133
BDM	66	70	136
Total	133	136	*269*

Note: BDM stands for the BeckerDeGrootMarschak mechanism; FPA stands for the First Price auction; HCL (LCL) stands for the High (Low) cognitive load treatment.

interactive, and included examples for each type of task that would appear in the experiment (see Experimental Instructions section in the Electronic Supplementary Material), with the exception of the valuation task where a practice and examples were not provided. Subjects were told that they will be given the opportunity to sell a card and earn more money but that the rules of selling the card will be given in detail when they reach the respective stage. We wanted to avoid subjects thinking through the task by the time they are actually confronted with it, so that if the cognitive load treatment has any meaningful effect we would be able to observe it. If the mechanics of the valuation task were known before subjects were exposed to the cognitive load treatment, then the treatment would be rendered ineffective by construction.

Subjects were instructed to raise their hand and ask any questions in private and that the experimenter (one of the authors) would then share his answer with the whole group. Subjects received a show-up fee of €3 and a fee of €3 for completing the experiment so that each subject would receive €6 with certainty upon successful completion of the experiment, which lasted about an hour. They could also earn additional money during the experiment from one of the randomly drawn tasks, and so the average of total payouts was €10.6 (S.D.=3.07, min=6, max=14).

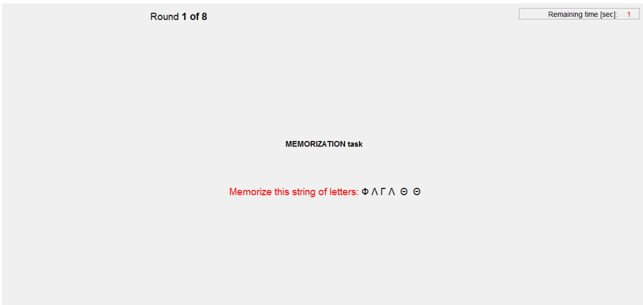
In total, subjects played 8 periods and in every period they went through one of the following decision tasks: 1) a valuation task, 2) an arithmetic (addition) task, 3) an arithmetic (multiplication) task, and 4) a click-a-button task. The valuation task was repeated twice as in Bull et al. (2019) and Cason and Plott (2014) since we intended to closely follow their design. Every other decision task was also repeated twice to match the repetitions of the valuation task. Subjects were not provided with any kind of feedback between periods for any of the tasks. Figure 1 shows sample screen shots illustrating the various decision tasks which are described in the next subsections in more detail.

2.1 Cognitive load manipulation

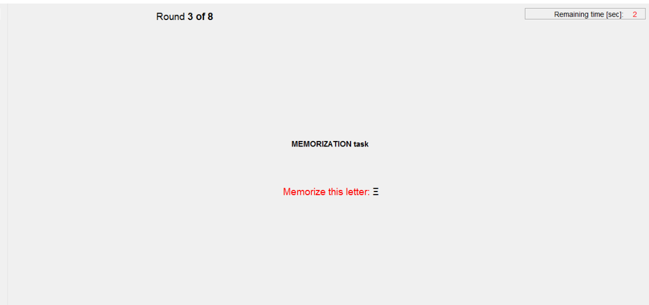
Cognitive load was manipulated by means of an incentivized *string* memorization task. While *number* memorization tasks have often been used in the literature (Benjamin et al., 2013; Deck and Jahedi, 2015; Drichoutis and Nayga, 2017), we used letters instead of numbers to avoid potential anchoring effects in the valuation task (Furnham and Boo, 2011). Specifically, in each period and just before the main decision task, a letter or a string of six letters appeared for four seconds on the participant's computer screen (see Figure 1a and 1b for sample screen shots). Subjects were then asked to keep this letter/string in their memory and recall it after the main decision task (see Figure 1h for a sample screen shot). If they recalled (typed) the letter/string correctly within a time limit of ten seconds, their memorization payoff for the period was €8. Otherwise it was €0. Subjects in the high cognitive load (HCL) treatment were shown six-letter strings while subjects in the low cognitive load (LCL) treatment were shown one letter to memorize. Letters (and strings) where drawn randomly in each period and independently from other subjects.

Figure 1: Sample screen shots of various tasks

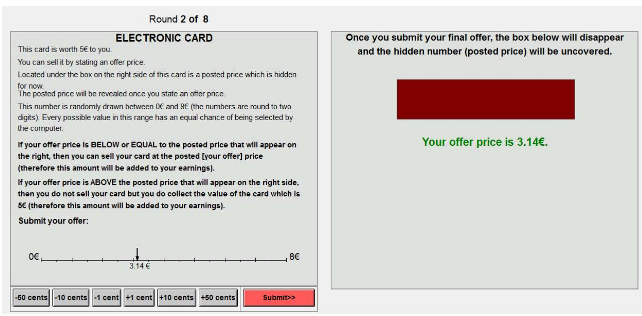
(a) Memorize string in the HCL treatment



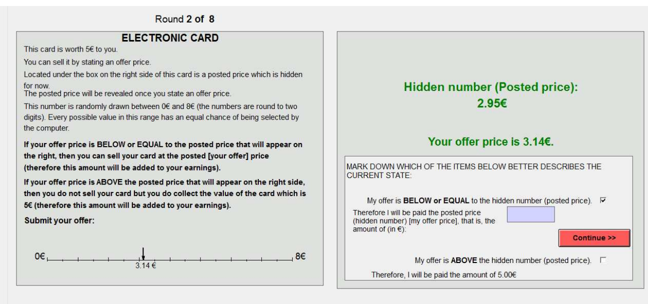
(b) Memorize letter in the LCL treatment



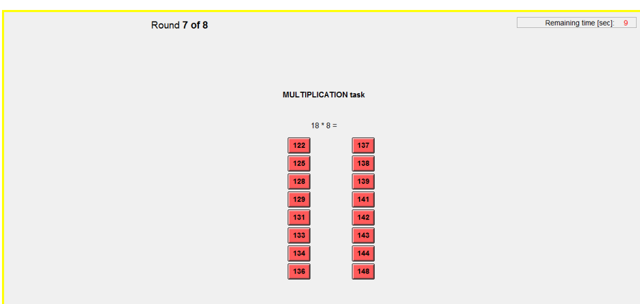
(c) Valuation task (selection of bid)



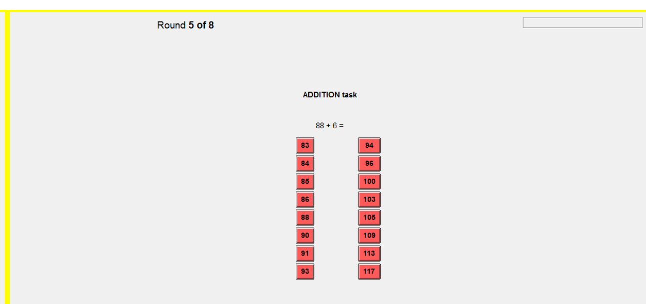
(d) Valuation task (posted price revealed)



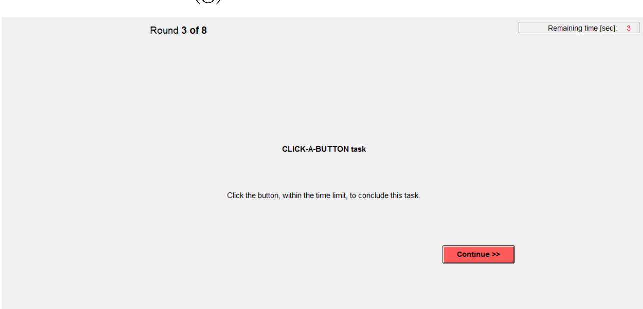
(e) Multiplication task



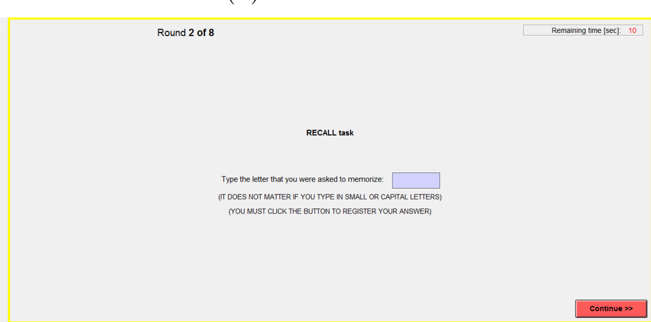
(f) Addition task



(g) Click-a-button task



(h) Recall number



2.2 The valuation task

Given that both Cason and Plott (2014) and Bull et al. (2019) used paper and pencil experiments, we tried to simulate their paper task as close as possible with a computer interface. In Cason and Plott (2014) and Bull et al. (2019), subjects were given a card with a front and a back side. Subjects would first state their offer for selling the card worth \$2 to them in the front side of the card. They were then instructed to turn around the card only after writing down a price and removing a tape to reveal the posted price.

To simulate this paper card, we showed on the computer screen two equally sized boxes, one on the left side of the screen (simulating the front side of the card) and one on the right side of the screen (simulating the right side of the card). Figures 1c and 1d show example screens of the valuation task. The left side box would contain the instructions (we used similar wording to Cason and Plott (2014) and Bull et al. (2019)) for selling the card as well as a radioline that subjects could click on in order to state their offer price. The radioline was anchored from €0 to €8 and was complemented with a series of buttons that could be used to add/subtract amounts of money to make more refined choices.

The right side of the card had a rectangle colored area (simulating the tape) that covered the randomly drawn price. Once subjects submitted and confirmed their offer, the buttons for making refined offers would disappear and the radioline was effectively inaccessible, preventing subjects from modifying their bid. At the same time, the rectangle area would disappear to reveal the posted price. After the posted price was revealed, subjects were asked to answer two questions taken verbatim from Cason and Plott (2014) and Bull et al. (2019) that reveal subjects' perception of the valuation task.

Our choice to scale up the induced value from \$2 to €5 was dictated by the fact that incentives in the main decision task had to be proportional to the certain payoff of subjects' participation. Cason and Plott (2014) and Bull et al. (2019) did not pay show up or participation fees because they did their valuation task to introductory classes with students. Our need to measure cognitive ability and precisely control the cognitive load procedure made the use of a computer laboratory environment necessary. The randomly drawn price was selected to be in the $[0, 8]$ interval and subjects were made aware of this.

The instructions between the BDM mechanism and the FPA were similar with the only difference being in this phrase: Subjects assigned to the BDM mechanism were told '...you can sell your card at *the posted* price' while subjects at the FP auction were told '...you can sell your card at *your offer* price'. If we denote the offer price chosen by the subject as b and the randomly drawn posted price as p, then the choice of an induced value of $IV = 5$ and an interval of posted prices of $[0, 8]$ determines the optimal offers as follows: i) in the BDM mechanism the expected payoff is $E[\pi] = IV \times \text{Prob}(b > p) + E(p|p > b) \times \text{Prob}(p > b)$ which simplifies to

$E[\pi] = 5\times b/8+(\frac{b+8}{2})\times(1-b/8)$. Therefore, expected payoff is maximized when $\partial E[\pi]/\partial b = 0$ or when $b_{BDM} = 5$, ii) in the FPA the expected payoff is $E[\pi] = IV \times \text{Prob}(b > p) + b \times \text{Prob}(p > b)$ which simplifies to $E[\pi] = 5 \times\ b/8 + b \times (1 - b/8)$. Therefore, expected payoff is maximized when $\partial E[\pi]/\partial b = 0$ or when $b_{FPA} = 6.5$.

2.3 Arithmetic and click-a-button tasks

We used the manipulation checks of Drichoutis and Nayga (2017) in order to identify whether the letter/string memorization task actually manipulates cognitive load. The tasks were meant to differ in terms of difficulty in order to assess the severity of the manipulation on decision making. Similar to Drichoutis and Nayga (2017) and Deck and Jahedi (2015), in the multiplication arithmetic task, subjects had to multiply two numbers. In the addition arithmetic task, subjects had to add two numbers.[2] Subjects had to indicate their answer by clicking the right choice from a list of randomly determined possible choices that were shown in two columns in an ordered manner; i.e., from low values to high values (see Figures 1e and 1f). The correct answer was set randomly to one of the buttons. In the click-a-button task, subjects simply had to click a button (see Figure 1g). The arithmetic and click-a-button tasks were set with a time limit of 11 seconds after which subjects would be forced out if they had not made a decision.

2.4 Cognitive ability

A cognitive load manipulation might have a differential effect on subjects with varying levels of working-memory capacity. Since working-memory capacity has been shown to be strongly correlated with general cognitive ability (Colom et al., 2004; Gray et al., 2003), we first measured the cognitive ability of all subjects using the Raven's Standard Progressive Matrices (RSPM) test which is used to assess mental ability associated with abstract reasoning and is considered a nonverbal estimate of fluid intelligence (Gray and Thompson, 2004). The RSPM test consists of 60 items and it took subjects on average 20 minutes to complete. Subjects were not provided with any feedback regarding their performance in the RSPM test. The RSPM test allows us to sum correct responses and form a measure of cognitive ability that we can then use to assess the effect of working-memory capacity on behavioral tasks' performance.

2.5 Payoffs and payments

Participants were paid for one randomly drawn period (out of 8 periods) and for only one of the (randomly determined) tasks in the randomly selected period (i.e., either the memorization

[2] In the multiplication arithmetic task, subjects had to multiply a one-digit integer $m_1 \sim U\{5, \ldots, 9\}$ and a two-digit integer $m_2 \sim U\{13, \ldots, 19\}$. In the addition arithmetic task, subjects had to add a one-digit integer $a_1 \sim U\{1, \ldots, 9\}$ and a two-digit integer $a_2 \sim U\{11, \ldots, 99\}$.

task or the decision task; depending on the period that was randomly drawn, the decision task could be either the valuation task or the addition task or the multiplication task or the click-a-button task). This was clearly explained beforehand in the instructions (see Experimental Instructions - Screen 2 in the Electronic Supplementary Material).

Similar to Deck and Jahedi (2015) and Drichoutis and Nayga (2017), we set the payoff associated with memorization higher than the payoff for the decision tasks, so that participants (even the ones with limited working memory capacity) would devote their attention to memorization and would not skip this task. Subjects would earn €8 if they recalled the letter/string correctly and €5 for a correct answer in either the multiplication task, the addition task or the click-a-button task. Our payoff scheme increases the likelihood that the cognitive load manipulation would be effective. In the valuation task, subjects could earn €5 (the Induced Value of the card) if their bid was higher than the randomly drawn price; they could also earn the posted price (their offer price) if their offer was lower than the posted price under the BDM mechanism (FPA).

3 Results

Before we present our results, it is useful to check whether our data show particular imbalance in terms of the observable characteristics of our subjects. While many researchers use statistical tests to check for balance of observable characteristics between treatments, the literature points to some pitfalls of this procedure (e.g., Briz et al., 2017; Deaton and Cartwright, 2016; Ho et al., 2007; Moher et al., 2010; Mutz and Pemantle, 2015). Following this literature, we report in Table 2 normalized differences (Imbens and Rubin, 2016; Imbens and Wooldridge, 2009) for all pairwise comparisons of treatment cells. For each subpanel of the table, we report in the lower diagonal normalized differences in location; i.e., normalized differences in means: $|\bar{x_1} - \bar{x_2}|/\sqrt{(s_1^2 + s_2^2)/2}$ where $\bar{x_j}$ and s_j^2 $(j = 1, 2)$ are the group means and variances, respectively. The upper diagonal parts report measures of differences in dispersion: $ln(s_1/s_2)$ (Imbens and Rubin, 2016). The dispersion difference measure indicates smaller differences in dispersion when its value is closer to zero. Cochran and Rubin's (1973) rule of thumb is that the normalized difference in location should be less than 0.25. Combining information from these two measures can give us some indication whether some characteristics systematically appear as particularly unbalanced. For example, age appears to have a bit larger normalized difference in means for a few cells of the table but the corresponding differences in dispersion are not problematic. More importantly, the Raven score variable which is important for our analysis, is not unbalanced.

Sections 3.1 and 3.2 establish that the memorization task was difficult to perform and that the memorization task affects tasks where reasoning is required (this is our manipulation check). The reader can directly skip to Section 3.3 where results for the bidding behavior are presented.

Table 2: Pairwise normalized differences between treatment cells for observable characteristics

			FPA		BDM	
			LCL	HCL	LCL	HCL
Gender	FPA	LCL		0.075	0.107	0.053
		HCL	0.272		0.031	-0.022
	BDM	LCL	0.340	0.067		-0.053
		HCL	0.216	0.055	0.122	
Age	FPA	LCL		-0.096	-0.001	-0.087
		HCL	0.541		0.095	0.008
	BDM	LCL	0.479	0.085		-0.086
		HCL	0.217	0.312	0.241	
Household size	FPA	LCL		0.304	-0.252	0.009
		HCL	0.185		-0.556	-0.295
	BDM	LCL	0.185	0.358		0.261
		HCL	0.149	0.016	0.315	
Income	FPA	LCL		-0.041	-0.104	-0.006
		HCL	0.076		-0.063	0.036
	BDM	LCL	0.095	0.021		0.099
		HCL	0.056	0.130	0.148	
Raven score	FPA	LCL		0.378	0.099	-0.092
		HCL	0.248		-0.279	-0.469
	BDM	LCL	0.236	0.015		-0.191
		HCL	0.077	0.145	0.144	

Notes: HCL (LCL) stands for the high (low) cognitive load treatment. For each subpanel of this table, the lower diagonal part shows normalized differences in location whereas the upper diagonal part shows normalized differences in dispersion.

3.1 Difficulty of the memorization task

We next explore whether the memorization task was difficult enough to have any meaningful effects. The top panel in Table 3 shows the frequency of correctly recalling the letter/string overall as well as after each task. It is obvious that success rate of recalling the string is significantly lower under cognitive load. Subjects are able to correctly recall the string one third of the times under load while they are successful almost every time when they just have to memorize a single letter in the low cognitive load treatment. A χ^2 test rejects the null for all rows of the top panel of Table 3 (p-value < 0.001) indicating that memorizing and correctly recalling a six letter string was significantly more difficult than memorizing and correctly recalling just one letter.

The difficulty of recalling the letter/string also exhibits some variation between tasks. For example, under low cognitive load, subjects almost perfectly recall the letter after the Click-a-button task (99.25%) while they fail to do so after the BDM/FPA (valuation) task (93.98%). A Fisher's exact test rejects the null of equality of success rates between tasks (p-value $<$ 0.003). Similarly, when under high cognitive load, subjects perform significantly worse after the multiplication task (22.79%) than after the Click-a-button task (45.22%). A χ^2 test rejects the null of equality of success rates between tasks (p-value < 0.001) under the HCL treatment.

Table 3: Success rate in the recall task

			HCL	LCL
Success rate	Combined over all tasks		36.21%	96.62%
	After…	Multiplication	22.79%	95.49%
		Addition	42.65%	97.74%
		Click-a-button	45.22%	99.25%
		Valuation task	34.19%	93.98%
Success rate		All 6 letters	36.21%	-
		5 letters	9.01%	-
		4 letters	9.38%	-
		3 letters or less	14.98%	-
		Did not submit anything	30.42%	-

Notes: HCL (LCL) stands for the high (low) cognitive load treatment. Differences between treatments (HCL vs. LCL) are statistically significant (p-value < 0.001) for all rows of the top panel of the table based on a χ^2 test.

The lower panel in Table 3 breaks down the percentage of successful recalls based on the count of letters subjects correctly typed (for example, if a subject was required to memorize '$M\Lambda KTM\Pi$' but then typed '$M\Lambda KT\Sigma\Lambda$', then this subject would have correctly typed four out of six letters). Table 3 shows that 54.60% of the times subjects correctly recalled 4 or more letters (= 36.21% + 9.01% + 9.38%) while 14.98% of the times subjects typed 3 letters or less correctly. 30.42% of the times (331 times out of 2152), subjects did not submit anything at all. This number is disproportionately shared between the Valuation task (35.65%) and the Multiplication task (31.42%) (versus 16.92% for the Addition task and 16.01% for the Click-a-button task), indicating that after a more demanding task it is more likely that subjects did not submit anything.[3]

In the Electronic Supplementary Material we present additional analysis where we estimate a logit model for the success/failure of recalling the memorized letter/string. This analysis supports the results above and generally establishes that the memorization task was indeed difficult to perform.

3.2 Manipulation checks

This section examines whether the memorization task was successful in loading the working memory of subjects so that it will result in worse outcomes in tasks where reasoning is required. Table 4 compares the success rates in each of the multiplication, addition, and click-a-button decision tasks. As shown, the memorization task successfully reduced success rates in the multiplication and addition tasks but had a negligible impact on the click-a-button. The reduction in success rates was larger in the multiplication task than in the addition task.[4] A one sided proportions test indicates that success rate in each decision task is lower under cognitive load, results in rejection of the null (for $\alpha = 10\%$) for the multiplication and addition tasks but not for the click-a-button task.

Table 4: Success rate in decision tasks

	HCL	LCL	p-value
Multiplication	44.85%	51.50%	0.061
Addition	84.92%	89.10%	0.075
Click-a-button	98.53%	99.25%	0.214

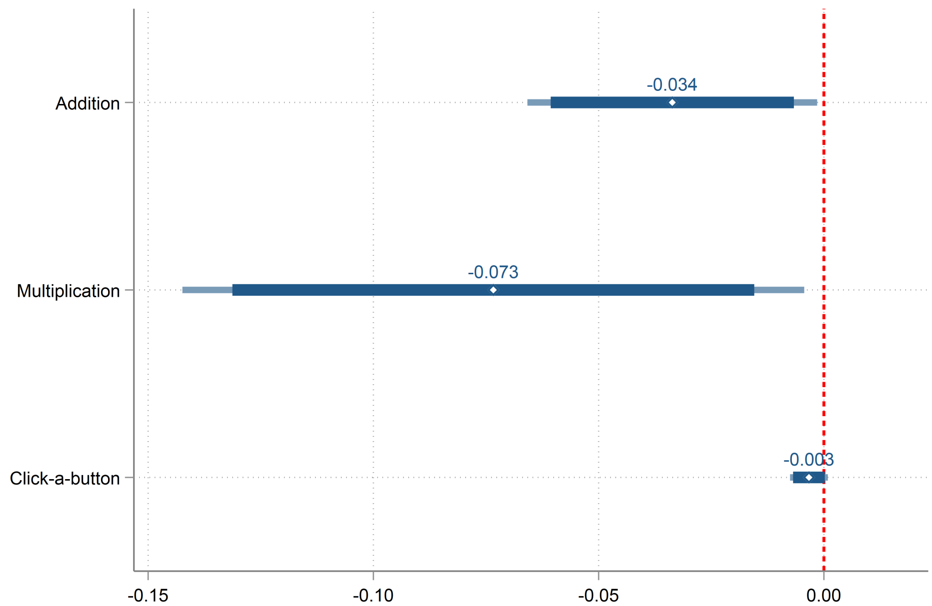
Notes: HCL (LCL) stands for the high (low) cognitive load treatment. Last column shows p-values comparing the two treatment for all rows of the table based on a one-sided proportions test.

In the Electronic Supplementary Material we present additional analysis where we run a

[3]On further reflection, subjects not submitting a string (i.e., submitting an empty field) could be due to one of the following reasons. Either a subject decides not to put much effort in recalling the string and clicks the button to proceed to the next screen or the subject puts some effort in recalling the string until time is up and does not have time to click on submit. We find that 28 choices (8.46% of choices that did not submit a string in the high cognitive load treatment) coming from 13 subjects submit an empty field before the time expires. This might be an indication that the vast majority of choices (91.54%) from subjects that did not submit a string (but could have typed a string and were just not fast enough to submit it) were putting some effort in the recall task.

[4]Table A8 in the Electronic Supplementary Material compares the reduction in success rate of correctly solving each of the three manipulation tasks with other studies that use number memorization techniques instead of letter memorization.

Figure 2: Marginal effects from Logit model (1) in Table A2 (with 95% and 90% CI)



logit regression of success/failure at the decision tasks, pooling data together from the three tasks (standard errors are clustered at the individual level) in order to econometrically control for the influence of observable characteristics and to explore the joint influence of the treatment variable and decision tasks.

Figure 2 graphs the marginal effects for cognitive load for the three decision tasks from the logit model (shown in the Electronic Supplementary Material). As shown, subjects are 7.3% less likely to correctly solve the multiplication task and 3.4% less likely to correctly solve the addition task when under cognitive load. On the other hand, cognitive load does not have a significant effect on the click-a-button task. Moreover, the positive effect of the Period variable in Table A2 indicates improved performance in these decision tasks as the experiment progresses. Note that the Raven score has a positive and statistically significant effect, indicating that subjects of higher cognitive ability are able to perform better in the decision tasks.

All in all, the results presented in this section show that the treatment was effective in inducing the desired effect according to our manipulation check. A significant treatment effect shows up even in the task where low reasoning is required (addition task) but not in a task where no reasoning is required (click-a-button task). Furthermore, we found that the effect increases in magnitude in a task involving high reasoning such as the multiplication task.

3.3 Descriptive analysis of bids

We first start by analyzing bidding behavior with simple graphs to gain some initial insights. Figure 3 depicts the histograms of bids per treatment in widths of 20 cents. Bars with thick outer lines depict overall frequency of bids around 4% of optimal offers for the BDM (€5±0.20 cents) and the FPA (€6.5±0.26 cents). The BDM treatment under low cognitive load (Figure 3a) echoes the results of Cason and Plott (2014) and Bull et al. (2019) in the sense that the BDM does not provide reliable measures of preferences.[5] In this treatment, the percent of offers within a range of 4% around the price of €6.5 (the optimal offer of the FPA) is barely 2.27%, indicating a very low proportion of subjects that potentially suffer from a FPA-GFM (i.e., they misperceive the BDM mechanism for a FPA).

Figure 3b indicates that subjects that bid in the BDM mechanism under high cognitive load show a somewhat different behavior. The proportion of bids around the offer of €5 drops to 5% from 15.15% (a 67% decrease; statistically significant according to a proportions test, p-value= 0.005) while the proportion of offers around the €6.5 offer increases from 2.27% to 10.71% (a 371.8% increase; statistically significant according to a proportions test, p-value= 0.005). Given that optimal offers around €6.5 become much higher, this might be an indication that the HCL treatment causes some FPA-GFM.

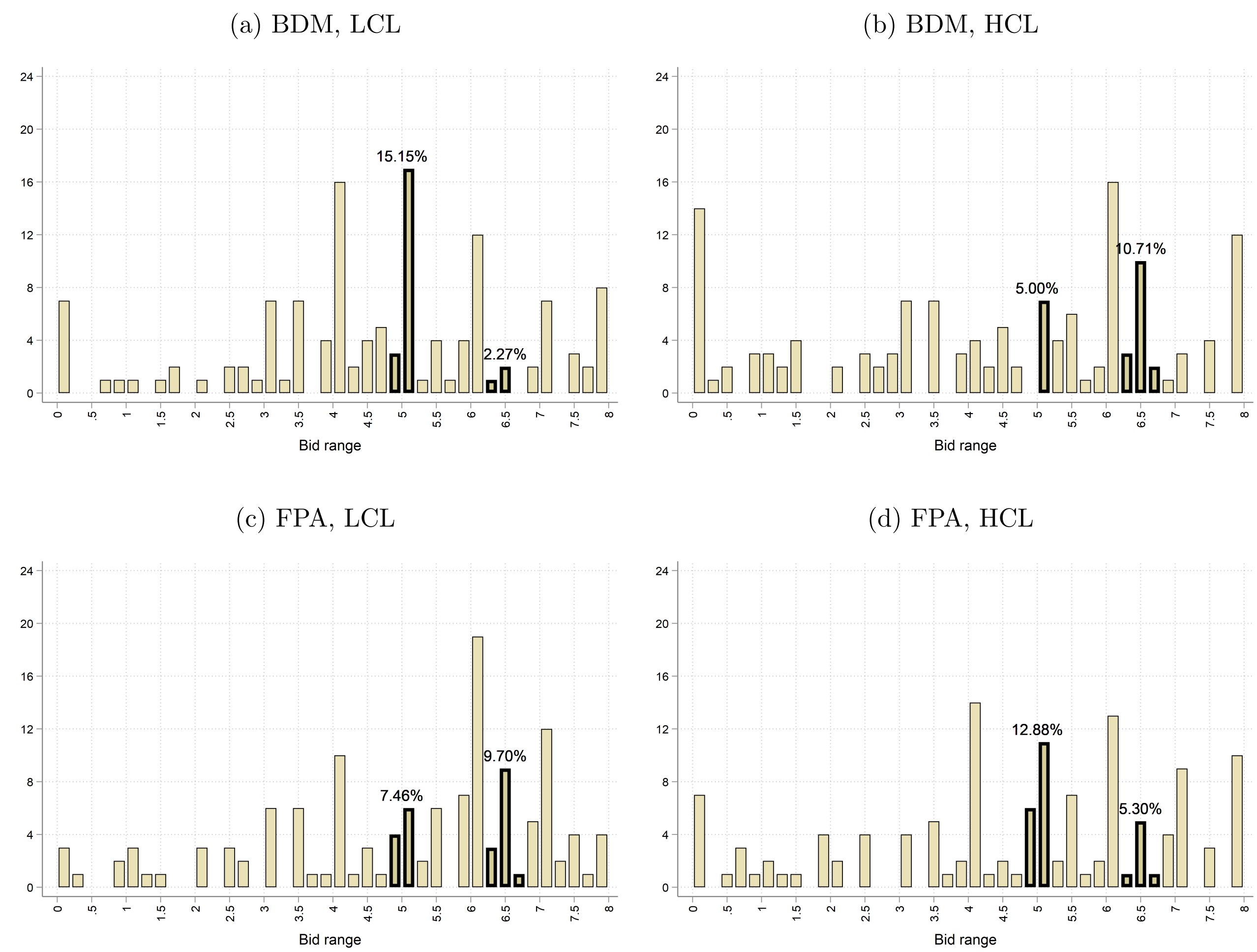
Examining behavior under the FPA is a useful comparison. Subjects in the FPA under low cognitive load (Figure 3c) exhibit a similar pattern in terms of proportion of offers around the €5 (not statistically significant according to a proportions test, p-value= 0.398) and the €6.5 bid with the BDM mechanism under high cognitive load (differences are not statistically significant according to a proportions test, p-value= 0.782). This result further reinforces the intuition that high cognitive load induces a FPA-GFM.

The last figure (Figure 3d) shows that in contrast to the FPA under LCL, applying high cognitive load in the FPA slightly increases offers around €5 and decreases offers around €6.5, although none of the differences is statistically significant (p-value= 0.144 for the proportion around €5 and p-value=0.174 for the proportion around €6.5). Moreover, as we show momentarily, the FPA-LCL and FPA-HCL bid distributions are indistinguishable in statistical terms.

Figure 4 shows kernel density estimates by treatment as well as the results of a Kolmogorov-Smirnov test comparing the bid distributions. Eye balling the graphs and the Kolmogorov-Smirnov tests support our discussion above: the bid distribution under the BDM mechanism is statistically different than the bid distribution under the FPA when subjects are exposed to

[5]In Bull et al. (2019) 7.9% of subjects make an optimal offer within 5 cents of the $2 induced value in Period 1 (16.7% in Cason and Plott (2014)) and 13.5% do so in Period 2 (31.1% in Cason and Plott (2014)). The proportion of subjects in our experiment bidding within 5 cents of the optimal offer in the BDM are similar to Bull et al. (2019): 7.58% bid within 5 cents of the optimal offer in Period 1 and 16.67% in P2. Improvement of outcomes in Period 2 is consistent with previous studies.

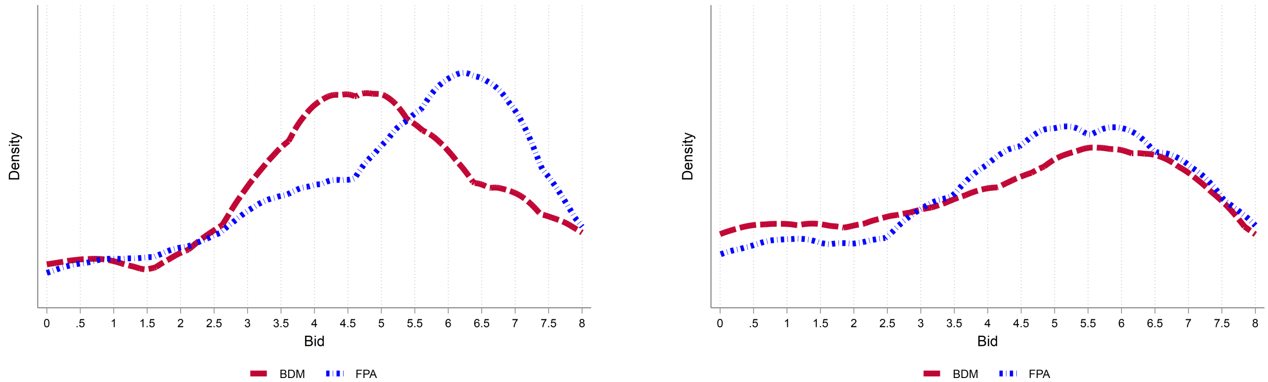
Figure 3: Histograms of bids in €0.2 widths



Notes: Bars with thick outer line indicate bids ±4% around the optimal offers of €5 and €6.5. HCL (LCL) stands for the high (low) cognitive load treatment.

Figure 4: Kernel density estimates of bids

(a) LCL (b) HCL



Note: Kolmogorov-Smirnov p-value = 0.003

Note: Kolmogorov-Smirnov p-value = 0.247

a LCL (Figure 4a) but the difference is muted under HCL (Figure 4b). This particular result indicates that a cognitive load manipulation may be able to turn on and off a FPA-GFM (if subjects really suffer from a FPA-GFM).

3.4 Optimal BDM mechanism offers or first-price game form misconceptions?

Although it is tempting to attribute differences in bid distributions that were described in the previous section to FPA-GFM, we might ask whether subjects' offers in the BDM mechanism are consistent with a model of optimal offers or with a first-price misconception model. To answer this question, we use maximum likelihood methods to fit our data in a model where subjects choose to maximize their expected expected payoff but make logit errors (Cason and Plott, 2014, also fitted similar models). Subjects' probability of submitting an offer b_j can be defined as:

$$\text{Prob}(\text{offer} = b_j) = \frac{e^{\lambda E[\pi|b_j]}}{\sum_n^{k=1} e^{\lambda E[\pi|b_k]}} \tag{1}$$

In this model the λ term bounds the cases where subjects are insensitive to differences in expected payoffs ($\lambda = 0$) or where subjects choose the offer that maximizes their expected payoff with the highest probability ($\lambda \to \infty$). A higher level of λ indicates a better fit, requiring less noise to characterize subject's choices according to that particular model.

In Equation 1, if we use $E^{opt}[\pi] = IV \times \text{Prob}(b > p) + E(p|p > b) \times \text{Prob}(b < p)$ then λ characterizes the optimal offers model of no misconceptions under the BDM mechanism. On the other hand, if we use $E^{gfm}[\pi] = IV \times \text{Prob}(b > p) + b \times \text{Prob}(b < p)$ then the estimated λ

characterizes the first price misconception model under the BDM mechanism. We then define the log-likelihood function as:

$$\ln L^m(\lambda; y_i) = \sum_i \ln \frac{y_i e^{\lambda E^m[\pi|b_j]}}{\sum_n^{k=1} e^{\lambda E^m[\pi|b_k]}} \tag{2}$$

where m stands for the optimal model ($m = opt$) or the FPA-GFM model ($m = gfm$) using the corresponding expected payoff expressions and y_i is an indicator that the offer is b_j.[6]

With risk averse bidders, optimal offers in a FPA deviate from the risk-neutral offer price of €6.5. With varying levels of risk, the optimal offer will be different for each subject. To show this, assume a Constant Relative Risk Aversion utility function $U(x) = \frac{x^{1-r}}{1-r}$ (where r is the coefficient or relative risk aversion). We can write the expected utility of a bidder submitting an offer b as:

$$EU(b; r, p) = \frac{IV^{1-r}}{1-r} \times \text{Prob}(b > p) + \frac{b^{1-r}}{1-r} \times \text{Prob}(p > b) \tag{3}$$

which can be rewritten as:

$$EU(b; r, \bar{p}) = \frac{IV^{1-r}}{1-r} \frac{b}{\bar{p}} + \frac{b^{1-r}}{1-r}(1 - \frac{b}{\bar{p}}) \tag{4}$$

Given an induced value of $IV = 5$, a maximum randomly drawn price of $\bar{p} = 8$ and setting $\frac{\partial EU}{\partial b} = 0$ gives:

$$5^{1-r} + (1-r)b^{-r}(8-b) - b^{1-r} = 0 \tag{5}$$

If we plug in $r = 0$ in Equation 5 we get $b = 6.5$, which is the optimal offer of the FPA under risk neutrality. We can then replace expected payoff in Equation 2 with Expected utility as defined in Equation 4:

$$\ln L^{RA}(\lambda; y_i) = \sum_i \ln \frac{y_i e^{\lambda EU[\pi|b_j]}}{\sum_n^{k=1} e^{\lambda EU[\pi|b_k]}} \tag{6}$$

Equation 6 bounds equation 2 as a special case for $r = 0$. Since we wanted to have the minimum amount of differences in the experimental design with respect to previous studies that examined FPA-GFM, in our experiment we did not elicit subjects' risk preferences. Nevertheless, we were able to match our data with data from a Holt and Laury (2002) risk preference elicitation task and a payoff-varying risk preference task (Kechagia and Drichoutis, 2017) (which is a scaled up version of the payoff varying task of Drichoutis and Lusk (2016)). The risk preferences

[6] Because the expression in the denominator of Equation 1 becomes extremely large when one uses the lowest possible division of 1 cent (i.e., the expression involves the summation of 801 summands), the maximum likelihood estimations are performed with y_i indicating offers being in a bin within 19 cents of b_j, which the lowest division that our estimation software would accommodate given the length of the expression involved. Cason and Plott (2014) use a similar strategy.

data come from an incentivized web survey that is being administered annually since 2017 to the student population of the university (administered online through the Qualtrics platform). Thus, we were able to match 156 subjects from the 2018 wave (the experiment described in this paper was conducted 2 months after the 2018 Qualtrics survey). We also matched an additional 24 subjects with risk preferences data from the 2017 wave and 23 more subjects from the 2019 wave. Consequently, we were able to match subjects from this experiment with risk preferences measures for the 75.46% of our sample.[7, 8]

There are two ways to integrate risk preferences in the analysis. First, we can estimate structural econometric models that allow us to calculate and predict a relative risk aversion (RRA) coefficient for each subject. This comes with the advantage that even for subjects that we do not have their risk preferences data, we can use the predictions for r based on the estimated parameters of the model. The Electronic Supplementary Data describes in detail how we went about the estimation and prediction of a RRA coefficient for each subject.[9] We can then use maximum likelihood methods to estimate the λ parameter in Equation 6. The statistical specification allows for the possibility of correlation between responses by the same subject.

Second, we can jointly estimate the r and λ parameters by stacking together choices made in the risk preferences task and the valuation task. The downside is that since we cannot match all subjects with the risk preferences data, we have to restrict our sample to the matched data. The Electronic Supplementary Material describes how we formulate the joint log-likelihood function in this case. Because results are similar when we use the predictions for r, we confine results from the joint estimation of r and λ in the Electronic Supplementary Material.

Table 5 shows the estimates from the sub-sample exposed to the BDM mechanism task. Model (1) shows the parameter estimates for λ from a model where the expected payoff expression is that of the optimal offer under the BDM mechanism. Model (2) shows the estimates when subjects are assumed to misconceive the expected utility expression in the BDM mechanism for that of the FPA, while model (3) presents the estimates from a mixture specification. In the mixture specification, we allow some choices to be consistent with the FPA-GFM model with probability π_{GFM} and consistent with the optimal offer model of BDM with probability

[7]In the web survey, subjects are given a €2 fixed fee for completing a questionnaire and a series of risk and time preferences tasks. A randomly drawn choice from the risk preferences task is added to their fixed fee and is then bank-transferred to subject's preferred bank account. Money are paid via the 'Pay a friend' service of the bank 'Eurobank' which allows transferring money to subject's preferred bank account without knowing subject's account number, only by using an email address or a mobile phone number. All transactions were ordered as soon as a task was completed and 88.44% of the ordered transactions went through. A dummy variable in the estimations controls for this fact.

[8]While the literature points to potential effects of cognitive load on risk version, Drichoutis and Nayga (2017) show that the meta-analytic effect from a set of four studies (Benjamin et al., 2013; Deck and Jahedi, 2015; Gerhardt et al., 2016; Olschewski et al., 2018) is actually a null effect.

[9]We find that subjects are, on average, characterized by risk aversion ($\bar{r} = 0.591$, min$= 0.258$, max$= 0.788$).

$1 - \pi_{GFM}$. Both models (2) and (3) are estimated with the restriction that $r = 0$ i.e., risk neutrality is imposed. Models (4) and (5) relax the risk neutrality assumption by using the values of r we predict from an independent estimation of risk preferences as described in the Electronic Supplementary Material.

Table 5: Maximum likelihood estimates of logit choice error parameter λ for optimal offers, first price auction misconception and mixture models

		Risk neutrality		Risk aversion	
	Optimal model (1)	FPA-GFM model (2)	Mixture model (3)	FPA-GFM model (4)	Mixture model (5)
λ					
Constant	1.202***	0.285***	1.105*	0.385**	1.170***
	(0.436)	(0.093)	(0.666)	(0.181)	(0.441)
HCL treatment	-1.140**	-0.219*	-1.039	-0.385**	-1.108**
	(0.533)	(0.126)	(0.671)	(0.181)	(0.537)
π_{GFM}					
Constant	-	-	0.084	-	0.088
	-	-	(0.294)	-	(0.256)
HCL treatment	-	-	0.915***	-	-0.088
	-	-	(0.294)	-	(0.256)
N (Subjects)			272 (136)		
Log-likelihood	-1011.551	-1012.372	-1010.748	-1015.772	-1011.465
AIC	2027.103	2028.744	2029.496	2035.545	2030.931
BIC	2034.314	2035.955	2043.919	2042.757	2045.354

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01.

Considering first the non-mixture models, it is obvious that the optimal offer model under the BDM exhibits a better fit across all models according to information criteria. Under low cognitive load, the λ parameter is estimated to be 1.202 and is significantly different than zero, indicating that subjects are not completely insensitive in how they choose their offer i.e., each bid value does not have the same probability of being offered. However, under high cognitive load, the overall value of the λ parameter goes down to 0.062 ($= 1.202 - 1.140$) which is not statistically significantly different than zero (p-value= 0.840). This is an indication that cognitive load causes a shift in the direction of making subjects choose all feasible offers with equal probability. The interpretation from the FPA-GFM model (model (2)) is similar, albeit the λ parameter is estimated to be significantly lower than the optimal offer model under BDM. Note that results from model (2) are very similar to results from model (4) where risk neutrality is relaxed.

The mixture model provides additional insights as per the model that best describes subjects' behavior. The probability that subjects' choice are better described by the FPA-GFM model under low cognitive load is not significantly different than zero ($\pi_{GFM} = 0.085$; p-value $= 0.774$) indicating that under low cognitive load, the optimal offer model for the BDM mechanism better describes subjects' behavior than the FPA-GFM model. In fact, this coefficient is very close to the one where risk neutrality is relaxed (model (5)). Although the high cognitive load treatment causes a significant increase in the estimated probability parameter that goes up from 0.084 to 0.999 in model (2) ($= 0.084 + 0.915$), indicating that the FPA-GFM model is the sole characterization of subjects behavior under HCL, the result is not robust to risk aversion.[10] From model (5) we infer that once risk neutrality is relaxed, the coefficient for the HCL treatment goes down to zero.

This is not a trivial change in inferences because a null coefficient for π_{GFM} suggests that the optimal model is the sole characterization of subjects bidding choices under the BDM mechanism treatment. To further explore the role of risk in the estimated probability π_{GFM} of a FPA-GFM, we estimated mixture models as in Table 5 but we assumed a value for r between 0 and 0.99 with steps of 0.01 (albeit we assumed that all subjects have the exact same r value). That is, we assumed varying levels of risk, ranging from risk neutraltity ($r = 0$) to high risk aversion ($r = 0.99$). Figure 5 graphs the estimated π_{GFM} for the LCL and HCL treatments. Consistent with Table 5, for risk neutrality and low levels of risk aversion ($r < 0.5$), the π_{GFM} for the LCL treatment is estimated close to zero while the π_{GFM} for the HCL treatment is estimated close to 1 giving rise to this dual characterization of subjects' bid choices when varying the cognitive load. However, for values of r greater than 0.5, then π_{GFM} is indistinguishable from zero for both the LCL and HCL treatments leaving the optimal model as the sole characterization of subjects' bidding choices under the BDM mechanism (π_{GFM} for the HCL treatment gradually shifts from 1 to 0 in the range of r values between 0.4 and 0.5).[11]

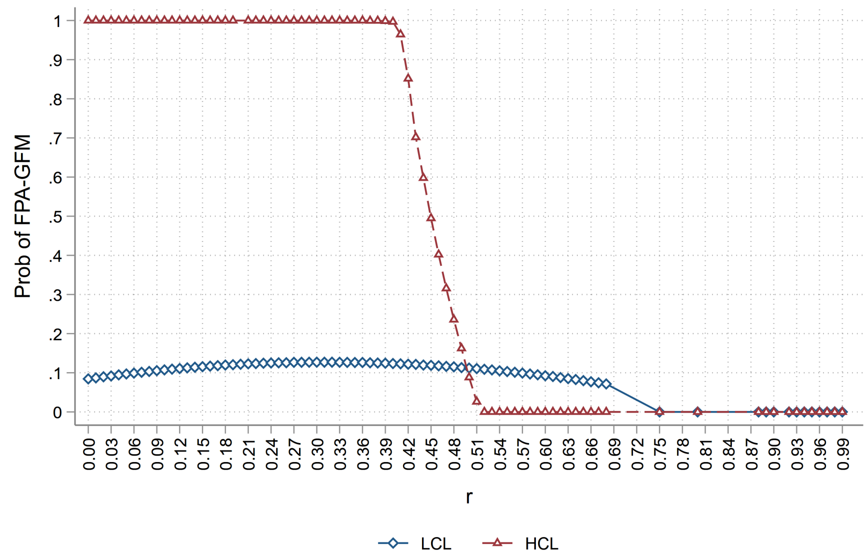
Overall, results from Table 5 indicate that the BDM optimal offer model fits our data better that every other model we tried, including a mixture specification where we let the data be determined by multiple models and estimate the probability of being consistent with one model or the other.

As a comparison, we also estimated a model where we assume that subjects in the BDM mechanism treatment suffer from FPA-GFM and pool the data together with the FPA treatment in order to estimate the λ values for each cell of our design. We use the interaction of the dummies of the treatments to allow us to estimate differential λ values for each cell. Results

[10] This probability is statistically different than zero (p-value$<$ 0.001) but not different than 1 (p-value$=$ 0.358).

[11] We note that this is not a general result which occurs when one integrates risk aversion in the analysis but rather a particular feature of our data. When we do the same exercise using Cason and Plott's (2014) data, we find that the estimated misconception probability of around 0.65 (as shown in column (1) of their Table 5) remains robust for values of r ranging from risk neutrality ($r = 0$) to high levels of risk aversion ($r = 0.99$). Figure A2 in the Electronic Supplementary Material depicts π_{GFM} for various levels of r using the Cason and Plott (2014) dataset and our estimation routines.

Figure 5: Probability of FPA-GFM by cognitive load and relative risk aversion coefficient



Note: LCL (HCL) stands for the low (high) cognitive load treatment.

are as follow (for the risk neutral case): $\lambda_{LCL}^{FPA} = 0.411$ (p-value < 0.001), $\lambda_{HCL}^{FPA} = 0.224$ (p-value $= 0.013$), $\lambda_{LCL}^{BDM} = 0.285$ (p-value $= 0.002$), $\lambda_{HCL}^{BDM} = 0.070$ (p-value $= 0.436$).[12] Note that inducing high cognitive load in the BDM treatment drives the value of λ down to zero (indicating that subjects choose offers with equal probability). This result echoes the estimated effect of the HCL treatment across all models of Table 5. The FPA treatment fits the data better than the BDM treatment (assuming that subjects in the BDM treatment think they are playing a FPA) since the estimated λ exhibits higher values than the BDM. A test that $\lambda_{LCL}^{FPA} = \lambda_{HCL}^{FPA} = \lambda_{LCL}^{BDM}$ does not reject the null ($\chi^2 = 1.67$, p-value $= 0.435$) but rejects the null at the 10% significance level that $\lambda_{LCL}^{FPA} = \lambda_{HCL}^{FPA} = \lambda_{LCL}^{BDM} = \lambda_{HCL}^{BDM}$ ($\chi^2 = 6.62$, p-value $= 0.085$).

3.5 Deviations from optimal bidding

In this section we explore how the treatments interact with cognitive ability (as proxied by the Raven score) to determine deviations from optimal offers when at the same time we control for observable characteristics. One way to explore this is by forming a measure of absolute deviations from the optimal offer for each valuation task. Since the optimal offers in the BDM mechanism and the FPA are different (€5 and €6.5, respectively), *similar* absolute deviations from the optimal offer will reflect *different* deviations of relative size in each valuation task. To account for this fact, we form a relative measure of deviations from the optimal

[12]Estimated values for λ when we relax the risk neutrality assumption are very close to the risk neutral estimates: $\lambda_{LCL}^{FPA} = 0.605$ (p-value $= 0.007$), $\lambda_{HCL}^{FPA} = 0.286$ (p-value $= 0.059$), $\lambda_{LCL}^{BDM} = 0.384$ (p-value $= 0.033$), $\lambda_{HCL}^{BDM} \approx 0$ (p-value $= 0.963$)

offer as: Relative bid deviation $= \left|\frac{\text{Bid}-\text{Optimal offer}}{\text{Optimal offer}}\right|$. We then run random effects regressions with clustered standard errors, with the relative bid deviations as the dependent variable. The independent variables include all the two-way interactions terms between the treatment variables and the Raven score variable to capture differential treatment effects that are mediated by subjects' level of cognitive ability. We also included a Period dummy, a set of demographic variables and a 'no-misconceptions' dummy. The no-misconceptions dummy was constructed based on the stated perceived payoff that subjects reported once the posted price was revealed in the valuation task when they were asked to state how much money they think they might receive based on the outcome of the task. The dummy is coded as 1 if subjects stated they will receive the posted price in the BDM mechanism or if they will receive the offer price in the FPA. Coefficient estimates from this model are reported in Table 6 (model (1) column). Since these coefficients are hard to interpret due to the presence of multiple interaction terms in the model, we present the marginal effect of the HCL treatment in graphical form in Figure 6a by valuation task and a range of values of cognitive ability. The range of values for cognitive ability are roughly the observed range of values in our sample.

Figure 6a shows that the effect of cognitive load under the FPA is not statistically significant for any value of the Raven score. On the other hand, subjects in the BDM treatment exhibit larger deviations when they are of low cognitive ability. For example, subjects in the lower range of the Raven score exhibit 17.5% larger deviations from the optimal offer under cognitive load but much lower relative deviations when they are of higher cognitive ability. In the extreme case where subjects are of very high cognitive ability, the effect of cognitive load is not statistically different from zero.[13]

Another way to analyze our data is by forming a dummy variable on whether a subject submitted a bid around ±1% of the optimal offer. Because of the binary nature of the dependent variable, we estimated a random effects Logit regression with clustered standard errors on the same set of independent variables as in model (1). Results are shown in model (2) of Table 6 albeit we focus on the graphical representation of the Marginal Effect of the HCL treatment in Figure 6b. Results are in accordance with results from model (1). The treatment effect of the HCL treatment is not distinguishable or different from zero in the FPA task for all levels of cognitive ability. However, in the BDM mechanism task subjects are significantly less likely to submit a bid around ±1% of the optimal offer. The effect is larger when subjects are of

[13] We can also relax the assumption of risk neutrality in the FPA by plugging in the estimated RRA coefficient r (the procedure to estimate r was described in the previous section) in equation 5, and solve the equation for b. Thus, the relative measure of bid deviation can also be constructed using an optimal offer that is conditioned on the level of risk aversion. Given subjects' risk aversion, we calculated the average optimal offer in the FPA to be €6.45 with a range of [6.43, 6.48] for the subjects of our sample. Since deviations from the risk neutral offer of 6.5 are so small, results using the risk averse optimal offer do not produce any meaningful changes in our results. Therefore, these results are confined to the Electronic Supplementary Material (see Figures A3, A5 and A7).

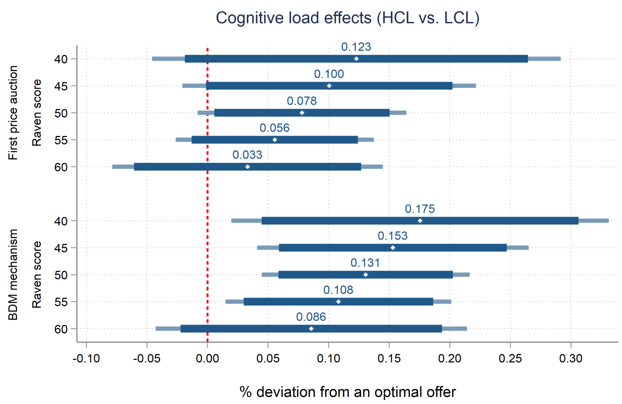
Table 6: Deviations from optimal offers: Random effects regression (Model 1), Random effects Logit regressions around ±1% of optimal offers (Model 2), Penalized ML Logit regression around ±1% of optimal offers (Model 3)

	% absolute bid deviation from optimal offer (1)	Deviation from optimal offer ±1% (2)	Deviation from optimal offer ±1% - PML (3)
Constant	1.034***	-6.440	-5.720
	(0.329)	(5.451)	(5.108)
HCL treatment	0.303	-8.902**	-7.772
	(0.312)	(4.491)	(5.097)
BDM	-0.798***	10.317**	9.620**
	(0.310)	(4.765)	(4.809)
Raven score	-0.012**	0.100	0.089
	(0.005)	(0.083)	(0.081)
HCL treatment × BDM	0.053	-0.314	-0.268
	(0.060)	(0.898)	(0.781)
BDM×Raven score	0.016***	-0.182**	-0.171*
	(0.006)	(0.091)	(0.091)
HCL treatment × Raven score	-0.004	0.154*	0.134
	(0.006)	(0.082)	(0.093)
No misconceptions	0.107***	-0.007	-0.001
	(0.024)	(0.382)	(0.370)
Period	-0.039**	0.064	0.059
	(0.015)	(0.363)	(0.348)
Demographics	Yes	Yes	Yes
$ln\sigma^2$		-1.787	
		(5.584)	
N	538	538	538
Log-likelihood	-	-120.255	-98.967
AIC	-	268.510	223.933
BIC	-	328.540	279.675

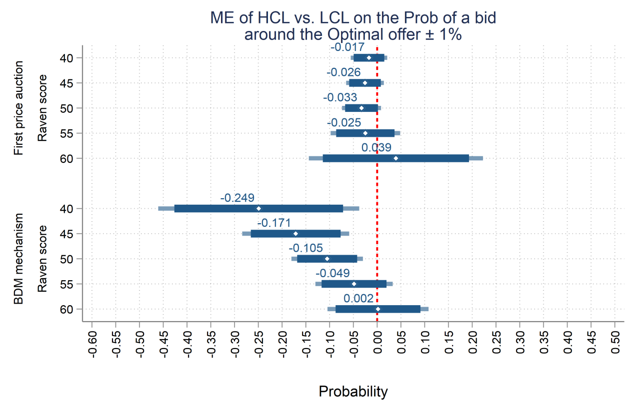
Notes: Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01. PML stands for penalized maximum likelihood.

Figure 6: Marginal effects of HCL treatment by Valuation mechanism and Raven score (with 95% and 90% confidence intervals)

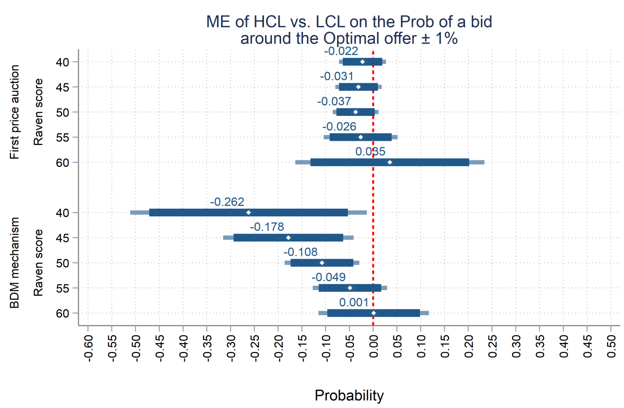
(a) ME from RE regression



(b) ME from RE Logit



(c) ME from PML Logit



Notes: ME stands for Marginal Effect; RE stands for Random Effects; PML stands for Penalized Maximum Likelihood

low cognitive ability (e.g., 24.9% less likely when they are of very low cognitive ability) but the effect becomes smaller for more cognitively able subjects. In fact, for subjects with very high cognitive ability, the effect of cognitive load is not statistically significantly different from zero.

Because we rather arbitrarily chose $\pm 1\%$ as the critical percent around an optimal offer to construct our dummy dependent variable, in the Electronic Supplementary Material we present additional figures that show that the marginal effect of the HCL treatment, as shown in Figure 6b, is robust to alternative definitions of the dependent variable as large as $\pm 8\%$ around the optimal offer (see Figure A4). For larger percentages we cannot detect statistically significant effects anymore.

Model (3) tackles a problem that might arise when one of the categories of the dependent variable is less frequent as in the case where we consider a $\pm 1\%$ range around the optimal offer. In this case, just 35 out of 538 cases are classified as optimal offers. The problem with this definition of an optimal offer is that maximum likelihood estimation (PMLE) suffers from small-sample bias and the bias depends on the number of cases in the less frequent of the two categories. Firth's (1993) method to address this problem is the Penalized maximum likelihood estimation (which is considered similar to King and Zeng's (2017) rare events Logit). Marginal effects from model (3) for the HCL treatment shown in Figure 6c show that results are robust even if we estimate the Logit model via the PMLE. The Electronic Supplementary Material shows that results for the marginal effect of the HCL treatment are robust to alternative definitions of the dependent variable as large as $\pm 8\%$ around the optimal offer (see Figure A6).

3.6 Perceived payoffs and instructions comprehension

Subjects were provided with detailed instructions about the bidding rules in the BDM mechanism and the FPA task. However, given that subjects had to read these instructions concurrently with performing the memorization task, it is likely that some of them decided to skip the instructions in order to focus on the string memorization task. Furthermore, their level of understanding may have been adversely affected due to the fact that string memorization was performed concurrently with the valuation task (although we would expect this to be affected more in the HCL treatment). Note, that all the other instructions about the rules of the experiment, payoffs, and trial rounds was shown before starting with the cognitive load treatment. The instructions for the BDM mechanism/FPA task were purposefully given concurrent with the memorization task for the reasons explained in Section 2.

In order to scrutinize subjects' attention to the instructions, we posed three questions related to the instructions at the end of the experiment by asking subjects to state: 1) how carefully they read the instructions in the BDM mechanism/FPA task on a scale from 1 to 5, anchored by 'not careful at all' and 'very careful' 2) how well they understood the instructions in the

BDM mechanism/FPA task on a scale from 1 to 5, anchored by 'not well at all' and 'very well' and 3) whether they decided to follow one of these mutually exclusive actions: a) to focus on the instructions and not memorize the letter/string, b) to focus on letter/string memorization and not pay attention to the instructions, c) to read the instructions while exerting effort in memorizing the letter/string.

Table 7 shows frequencies for each level of the Likert scales for the three questions mentioned above. In general, most subjects stated they carefully read and well comprehended the instructions. There is not much difference between the BDM/FPA task but there is a marked shift in the distribution of responses in the HCL treatment in terms of reading and comprehension of instructions. A Fisher's exact test rejects the null that responses in the 'carefully read instructions' question are spread equally across the treatment cells (p-value = 0.005) but does not reject the null in the comprehension of instructions question (p-value = 0.147). The statistical significant result for carefully reading the instructions is fully driven by the HCL treatment (which is to be expected just by eye-balling Table 7).[14]

Table 7: Comprehension of instructions in the BDM mechanism/FPA task

Question/Scale		LCL			HCL		
		FPA	BDM	*FPA+BDM*	FPA	BDM	*FPA+BDM*
Carefully read instructions for the BDM (FPA) task	1 = Not careful at all	1	0	*1*	0	0	*0*
	2	1	0	*1*	1	1	*2*
	3	5	3	*8*	14	17	*31*
	4	23	33	*56*	27	26	*53*
	5 = Very careful	37	30	*67*	24	26	*50*
How good did subject comprehend instructions in the BDM (FPA) task	1= Not well at all	2	0	*2*	1	1	*2*
	2	0	0	*0*	1	0	*1*
	3	5	6	*11*	6	12	*18*
	4	27	39	*66*	38	33	*71*
	5 = Very well	33	21	*54*	20	24	*44*
During the experiment, subject was ...							
1= ...focused on reading instructions for the BDM (FPA) task		4	3	*7*	12	14	*26*
2= ...focused on memorizing the string		1	0	*1*	6	7	*13*
3= ...paid attention in reading instructions for the BDM (FPA) task as well as memorizing the string		62	63	*125*	48	49	*97*

Notes: HCL (LCL) stands for the high (low) cognitive load treatment. Table shows frequencies.

The responses to the final question in this set of questions show that very few subjects decided to focus on either just reading the instructions for the valuation task or just memorizing the letter/string. Most subjects focused on both concurrent tasks, although there is a shift of

[14]Fisher's exact test p-value is < 0.001 when we compare the HCl vs. LCL treatments but is 0.716 when we compare the BDM vs. FPA. In terms of comprehension of instructions, binary comparisons of treatments do not show any statistically significant result either: Fisher's exact test p-value is 0.327 (0.402) when we compare the HCL vs. LCL (BDM vs. FPA) treatments.

responses under the HCL treatment indicating that more subjects decided to focus on one of the two tasks when under cognitive load. Fisher's exact test p-value is < 0.001 when we compare the HCL vs. LCL treatments and is close to 1 when we compare the BDM vs. FPA treatments.

Furthermore, since we recorded time responses of subjects in the valuation task (we recorded time until subjects finalize their bid and time until they exited the valuation task), we run random effects regressions (with clustered standard errors at the individual level) of time on the treatment dummies as well as on dummies for the levels of the instructions-related variables described above. Results are exhibited in Table 8. A few things are noteworthy in this table. First, it is evident that subjects in the second period spend significantly less time reading the instructions, submitting a bid, and exiting the valuation stage. For example, they spend about 53 seconds less in submitting a bid and about 75 seconds less to complete the valuation task. Moreover, more cognitively able subjects spend more time submitting a bid (which includes reading the instructions) and completing the valuation task. With respect to the instructions-related variables, only the dummies indicating carefully reading the instructions are significantly affecting time spent. For example, a subject stating she read the instructions 'very carefully' spent 39.5 seconds more reading the instructions and deciding a bid than a subject stating she was 'not careful at all' in reading the instructions.

To take into account attention to instructions, we rerun the main analysis described in Sections 3.4 and Section 3.5 using a restricted sample of subjects; i.e., the union of subjects that stated to carefully read and comprehend the instructions at a moderate level or better and those that stated they paid attention in both reading the instructions and memorizing the letter/string. This restricts our sample to 221 subjects (111 in the BDM treatment). Results are confined to the Electronic Supplementary Material (Table A10 and Figure A8) but can be briefly summarized as follows: a) the BDM optimal offer model is the best fitting model for the restricted sample while a mixture specification indicates that the BDM optimal offer model is the sole characterization of subjects' choices and b) the effect of cognitive load in the FPA treatment does not vary with cognitive ability, while subjects in the BDM treatment exhibit larger deviations from optimal offers (as well as they are more likely to deviate from optimal offers) when they are of low cognitive ability.

In addition to subjects' stated attention and comprehension to instructions at the valuation stage, we also asked subjects (after submitting their bid and seeing the posted price) to indicate: a) whether their bid is lower or greater than the posted price and b) if they stated that their bid is lower than the posted price, to write down how much they think they will be paid. Table 9 categorizes subjects' stated payoffs by valuation task and treatment. As can been seen, in most cases, subjects stated that they will either be paid the posted price (which is correct under the BDM mechanism) or their offer price (which is correct under the FPA task).[15] For eight cases

[15]For a handful of cases, subjects also stated that they thought they will be paid an amount equal to the

Table 8: Random effects regressions of time spent in valuation stage

	(1)		(2)	
	Time to submit a bid		Total time to exit the valuation stage	
Constant	19.123	(25.951)	69.205**	(34.468)
BDM	-0.746	(2.778)	3.256	(3.617)
HCL treatment	-1.858	(3.006)	-1.717	(3.938)
2nd Period	-53.354***	(2.286)	-75.550***	(2.904)
Raven score	0.856***	(0.252)	0.810**	(0.335)
Carefully read instructions				
2	10.713	(7.343)	36.775***	(14.232)
3	20.358**	(10.372)	37.825***	(12.260)
4	33.388***	(11.118)	54.090***	(13.870)
5 = Very careful	39.483***	(10.934)	61.387***	(13.569)
Instructions comprehension				
2	-33.712***	(10.362)	-4.632	(12.105)
3	-2.569	(10.083)	-5.437	(9.599)
4	-3.497	(10.598)	-4.293	(10.351)
5 = Very well	-5.963	(11.114)	-12.001	(10.791)
Focused on ...				
...reading instructions	3.481	(6.128)	5.021	(7.259)
...memorization task	-3.242	(4.728)	1.931	(9.863)
Male	-0.839	(3.110)	0.903	(4.043)
Age	-0.622	(0.941)	-1.955*	(1.170)
N			538	

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01. HCL (LCL) stands for the high (low) cognitive load treatment. Base categories are: Carefully read instructions: 1 = not at all; Instructions comprehension: 1 = not well at all; Focused on: reading instructions & memorizing the letter/string.

(7.08% of the BDM cases), subjects incorrectly stated that they will be paid their offer price under the BDM. These are the cases that are directly revealed to have a FPA-GFM and are equally spread among the two treatments. Similarly, there are only four cases in the FPA task that they incorrectly stated they will be paid the posted price and these are all cases under the high cognitive load treatment.

Table 9: Perceived payoffs by valuation task and treatment

Perceived payoff is ...	BDM		FPA	
	LCL	HCL	LCL	HCL
Offer price	4	4	*45*	*40*
Posted price	*50*	*46*	-	4
Posted price + induced value	-	2	-	-
Posted + Offer price	2	1	-	1
Posted - Offer price	-	-	-	4
Other	2	2	4	3
Total	58	55	49	52

Notes: HCL (LCL) stands for the high (low) cognitive load treatment. Table shows frequencies.

Table 10 tabulates subjects' perception about whether their bid was lower/greater than the posted price with whether the bid was actually lower/greater than the posted price as well as the mean perceived and actual payoff by treatment group. The upper right part and lower left part of the table indicate cases where subjects incorrectly stated their bid is lower or greater than the posted price. A comparison of perceived and actual payoff for those that correctly stated their bid to be smaller than the posted price shows that these are very close and no significant differences between treatments can be observed. Pulling together all responses and doing a t-test of whether actual payoffs are equal to perceived payoffs, we fail to reject the null of no difference (p-value = 0.632). We get similar non-statistically significant results if we do the test by treatment or by period.

We then rerun the main analysis described in Sections 3.4 and Section 3.5 further restricting our sample to those subjects that did not misperceive the magnitude of their bid with respect to the posted price. This reduces our sample to 208 subjects (104 in the BDM treatment). Results are shown in the Electronic Supplementary Material (Table A11 and Figure A9). None of our conclusions changes.

sum of the posted price and the induced value, or equal to the posted price plus the offer price, or equal to the posted price minus the offer price, or they stated an amount that cannot be assigned to one of the categories.

Table 10: Tabulation of subjects perception about their bid and actual bid (in relation to posted price) and average perceived payoffs (in relation to actual payoffs)

			Bid < Posted price			Bid > Posted price	
Subject perceives that ...	Treatments		N	Perceived Payoff	Actual Payoff	N	Actual Payoff
Bid < Posted price	HCL	BDM	55	5.53	5.45	*12*	*4.58*
		FPA	49	3.64	3.47	*4*	*3.06*
	LCL	BDM	55	6.08	6.23	*4*	*6.66*
		FPA	48	4.06	3.88	*2*	*1.56*
Bid > Posted price	HCL	BDM	*0*	-	-	73	5.00
		FPA	*3*	*4.27*	*5.00*	76	5.00
	LCL	BDM	*3*	*3.61*	*5.00*	70	5.00
		FPA	*1*	*8.00*	*5.00*	83	5.00

Notes: HCL (LCL) stands for the high (low) cognitive load treatment. The upper right part and lower left part of the table show cases for which subjects incorrectly state their bid is greater or lower than the posted price.

4 Discussion and conclusions

In this study, we focused on the increasingly important issue related to failure of game form recognition in the BDM mechanism as shown by Cason and Plott (2014). Building on the work of Bull et al. (2019) that added a First-Price Auction task on top to the BDM mechanism task as an additional test of the game form misconception, we examine whether limiting cognitive resources of subjects would exacerbate game form misconception in the BDM. We hypothesized that because the BDM mechanism and the FPA differ in their difficulty of understanding the rules, the BDM task would be more severely affected under higher cognitive load.

Our results suggest that although the BDM mechanism was far from accurate in revealing subjects' true preferences, we do not find that offers in this task are consistent with a FPA misconception model. This result remains robust even when we take into account in the analysis additional issues such as subjects' risk aversion, comprehension of instructions, and payoffs of the valuation task as well as subjects' involvement with the valuation task. Interestingly, what we do find is that when subjects are placed under high cognitive load, their decision process is equivalent to random choice of an offer since their choices are consistent with a model where subjects choose all offers with equal probability. This only occurs however, in the BDM mechanism task since subjects in the FPA do make offers consistent with maximizing their expected payoff.

Moreover, we find that the effect of cognitive load in the FPA treatment does not vary with cognitive ability. Subjects in the BDM treatment however, exhibit larger deviations from optimal offers and they are more likely to deviate from optimal offers when they are of low cognitive ability. Subjects with higher cognitive ability are unaffected by high cognitive load.

The FPA is also immune to the high cognitive load treatment, leaving the BDM as a questionable mechanism for accurate measurement of subjects' preferences.

We agree with Cason and Plott (2014) that choices cannot always be interpreted reliably as revealing preferences. However, with our data we do not find support that a FPA-misconception is occurring in the low cognitive load condition. By inducing a high cognitive load, we are giving the best chances for a FPA-game form misconception to occur, but we do not see this happening as well. Instead, we find that subjects are better characterized as choosing randomly from the set of all offer prices rather than mis-perceiving the BDM mechanism for a FPA. We agree with Bull et al. (2019) that subjects have poor optimizing skills, resulting in very noisy data that is especially pronounced in the BDM task under the HCL treatment.

One particular feature of our experiment is that everything was computerized and as such, the flow of the experiment, instructions and even the bidding process were likely easier than the paper and pencil experiments of Cason and Plott (2014) and Bull et al. (2019). It is possible that a paper and pencil administration of the experiments could induce some higher cognitive load than our LCL treatment; and if this is the case, their experiments could be seen perhaps as falling in between our LCL and HCL treatments in terms of cognitive resources depletion.

Our results suggest that choices made in environments that are cognitively demanding or in environments that can deplete cognitive resources are prone to eliciting randomness in choices rather than revealing preferences. However, our results also point to the fact that different mechanisms may be more or less susceptible to elicitation of random choices rather than preferences. If anything, our study builds on the accumulated literature pointing to problems with preference revelation using the BDM mechanism (Banerji and Gupta, 2014; Horowitz, 2006; Karni and Safra, 1987; Mazar et al., 2013; Rosato and Tymula, 2016; Urbancic, 2011; Vassilopoulos et al., 2018).

References

Banerji, A. and N. Gupta (2014). Detection, identification, and estimation of loss aversion: Evidence from an auction experiment. *American Economic Journal: Microeconomics 6*(1), 91–133.

Becker, G. M., M. H. Degroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science 9*(3), 226–232.

Benjamin, D. J., S. A. Brown, and J. M. Shapiro (2013). Who is 'behavioral'? Cognitive ability and anomalous preferences. *Journal of the European Economic Association 11*(6), 1231–1255.

Bigler, J. (2019, May 10). An update on first price auctions for Google Ad Manager. Available at https://perma.cc/U3ZC-NY79. Last accessed on November 15, 2019.

Brañas Garza, P. and J. Smith (2016). Cognitive abilities and economic behavior. *Journal of Behavioral and Experimental Economics 64*, 1–4.

Briz, T., A. C. Drichoutis, and R. M. Nayga Jr (2017). Randomization to treatment failure in experimental auctions: The value of data from training rounds. *Journal of Behavioral and Experimental Economics 71*, 56–66.

Bull, C., P. Courty, M. Doyon, and D. Rondeau (2019). Failure of the Becker-DeGroot-Marschak mechanism in inexperienced subjects: New tests of the game form misconception hypothesis. *Journal of Economic Behavior & Organization 159*, 235–253.

Busemeyer, J. R. and J. T. Townsend (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences 23*(3), 255–282.

Busemeyer, J. R. and J. T. Townsend (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review 100*(3), 432–459.

Cason, T. N. and C. R. Plott (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy 122*(6), 1235–1270.

Choi, S., S. Kariv, W. Müller, and D. Silverman (2014). Who is (more) rational? *American Economic Review 104*(6), 1518–50.

Clarke, K. A. (2003). Nonparametric model discrimination in international relations. *Journal of Conflict Resolution 47*(1), 72–93.

Cochran, W. G. and D. B. Rubin (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A 35*(4), 417–446.

Colom, R., I. Rebollo, A. Palacios, M. Juan-Espinosa, and P. C. Kyllonen (2004). Working memory is (almost) perfectly predicted by g. *Intelligence 32*(3), 277–296.

Deaton, A. and N. Cartwright (2016). Understanding and misunderstanding randomized controlled trials. *National Bureau of Economic Research Working Paper No. 22595*.

Deck, C. and S. Jahedi (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review 78*, 97–119.

Diggle, P. J., P. Heagerty, K.-Y. Liang, and S. L. Zeger (2002). *Analysis of Longitudinal Data* (2nd ed.). New York, USA: Oxford University Press Inc.

Drichoutis, A. C. and J. L. Lusk (2014). Judging statistical models of individual decision making under risk using in- and out-of-sample criteria. *PLoS ONE 9*(7), e102269.

Drichoutis, A. C. and J. L. Lusk (2016). What can multiple price lists really tell us about risk preferences? *Journal of Risk and Uncertainty 53*(2), 89–106.

Drichoutis, A. C. and J. Nayga, Rodolfo M. (2017). Economic rationality under cognitive load. *Munich Personal RePEc Archive No. 81735*.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika 80*(1), 27–38.

Furnham, A. and H. C. Boo (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics 40*(1), 35–42.

Gerhardt, H., G. P. Biele, H. R. Heekeren, and H. Uhlig (2016). Cognitive load increases risk aversion. *SFB 649 Discussion Paper 2016-011*.

Gray, J. R., C. F. Chabris, and T. S. Braver (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience 6*(3), 316–322.

Gray, J. R. and P. M. Thompson (2004). Neurobiology of intelligence: science and ethics. *Nature Reviews Neuroscience 5*(6), 471–482.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association 1*(1), 114–125.

Hassidim, A., D. Marciano, A. Romm, and R. I. Shorrer (2017). The mechanism is truthful, why aren't you? *American Economic Review 107*(5), 220–24.

Hey, J. D., G. Lotito, and A. Maffioletti (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of Risk and Uncertainty 41*(2), 81–111.

Ho, D. E., K. Imai, G. King, and E. A. Stuart (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis 15*(3), 199–236.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *The American Economic Review 92*(5), 1644–1655.

Horowitz, J. K. (2006). The Becker-DeGroot-Marschak mechanism is not necessarily incentive compatible, even for non-random goods. *Economics Letters 93*(1), 6–11.

Imbens, G. W. and D. B. Rubin (2016). *Causal Inference for Statistics, Social, and Biomedical Sciences, An introduction*. Cambridge and New York: Cambridge University Press.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Karni, E. and Z. Safra (1987). "Preference reversal" and the observability of preferences by experimental methods. *Econometrica 55*(3), 675–685.

Kechagia, V. and A. C. Drichoutis (2017). The effect of olfactory sensory cues on willingness to pay and choice under risk. *Journal of Behavioral and Experimental Economics 70*, 33–46.

King, G. and L. Zeng (2017). Logistic regression in rare events data. *Political Analysis 9*(2), 137–163.

Kupper, L. L. and K. B. Hafner (1989). How appropriate are popular sample size formulas? *The American Statistician 43*(2), 101–105.

Lee, J. Y., R. M. J. Nayga, C. Deck, and A. C. Drichoutis (2017). Cognitive ability and bidding behavior in second price auctions: An experimental study. *Munich Personal RePEc Archive No. 81495*.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review 107*(11), 3257–3287.

Liu, H. and T. Wu (2005). Sample size calculation and power analysis of time-averaged difference. *Journal of Modern Applied Statistical Methods 4*(2), 434–445.

Mazar, N., B. Koszegi, and D. Ariely (2013). True context-dependent preferences? The causes of market-dependent valuations. *Journal of Behavioral Decision Making*.

Moher, D., S. Hopewell, K. F. Schulz, V. Montori, P. C. Gotzsche, P. J. Devereaux, D. Elbourne, M. Egger, and D. G. Altman (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ 340*.

Mutz, D. C. and R. Pemantle (2015). Standards for experimental research: Encouraging a better understanding of experimental methods. *Journal of Experimental Political Science 2*(2), 192–215.

Norwood, B. F., M. C. Roberts, and J. L. Lusk (2004). Ranking crop yield models using out-of-sample likelihood functions. *American Journal of Agricultural Economics 86*(4), 1032–1043.

Olschewski, S., J. Rieskamp, and B. Scheibehenne (2018). Taxing cognitive capacities reduces choice consistency rather than preference: A model-based test. *Journal of Experimental Psychology: General 147*(4), 462–484.

Rosato, A. and A. A. Tymula (2016). Loss aversion and competition in Vickrey auctions: Money ain't no good. *The University of Sydney, Economics Working Paper Series 2016 - 09*.

Rustichini, A. (2015). The role of intelligence in economic decision making. *Current Opinion in Behavioral Sciences 5*, 32–36.

StataCorp (2013). *Stata user's guide release 13.* College Station, Texas, USA: Stata Press.

Urbancic, M. (2011). Testing distributional dependence in the Becker-DeGroot-Marschak mechanism. *Working paper, UC Berkeley*.

Vassilopoulos, A., A. C. Drichoutis, and R. M. Nayga Jr (2018). Loss aversion, expectations and anchoring in the BDM mechanism. *Munich Personal RePEc Archive No. 85635*.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*(2), 307–333.

Wilcox, N. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In J. C. Cox and G. W. Harrison (Eds.), *Research in Experimental Economics Vol 12: Risk Aversion in Experiments*, pp. 197–292. Bingley, UK: Emerald Group Publishing Limited.

Wilcox, N. (2015). Error and generalization in discrete choice under risk. *Economic Science Institute (Chapman University) Working paper*.

Wilcox, N. T. (2011). 'Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics 162*(1), 89–104.

Electronic Supplementary Material of

Game form recognition in preference elicitation, cognitive abilities and cognitive load

Andreas C. Drichoutis* and Rodolfo M. Nayga†

Experimental instructions

Instructions were provided in electronic form within the zTree environment. This is a translation of the original instructions written in Greek.

Screen 1

Welcome to our study!

With your presence here today, you have secured 3€. In addition, you will receive 3€ when you finish the study. That is, you will get 6€ for sure by completing today's study.

In addition, you can earn extra money depending on the decisions you make and therefore it is of great importance to read the instructions that follow carefully. If you have any questions during the survey, raise your hand and a researcher will come close to you.

It is very important not to communicate with any of the other participants.

VERY IMPORTANT: Please remove from your desk the following items: mobile phone, paper, notebook, pen, pencil.

*Corresponding author: Associate Professor, Department of Agricultural Economics & Rural Development, Agricultural University of Athens, Iera Odos 75, 11855, Greece, e-mail: adrihout@aua.gr.

†Distinguished Professor and Tyson Endowed Chair, Department of Agricultural Economics & Agribusiness, University of Arkansas, Fayetteville, AR 72701, USA, tel:+1-4795752299 e-mail: rnayga@uark.edu.

Screen 2

Instructions

In this study threre will be 8 rounds. In every round you will perform two tasks. The **first task** will be a memorization task. The **second task** will differ from round to round and will be described in the screens to follow.

Overall, you will complete 16 tasks and will make the respective decisions, One of these decisions will be randomly selected at the end of the study and you will get paid ONLY for this decision.

Given you don't know which task will be randomly selected, you need to treat each task seriously, since each task has exactly the same probability to be randomly selected and count toward your earnings.

Continue >>

Screen 3

(note: text in brackets indicates different text used in the high/low cognitive load treatments)

The MEMORIZATION task

In the beginning of every round you will be given a letter [string of letters] to memorize. The letter [string of letters] will remain visible for 3 seconds and then you will be given a **Second task**. As soon as you complete the second task, you will be asked to recall the letter [string of letters] you memorized in the beginning of the round. If you recall the letter [string of letters] correctly, your earnings from the memorization task will be **8€**. If you don't recall the letter [string of letters] correctly, your earnings from the memorization task will be **0€**.

In the next screens you will be given the chance to practice the memorization task. Practice does not count toward your earnings.

Continue >>

Screen 4

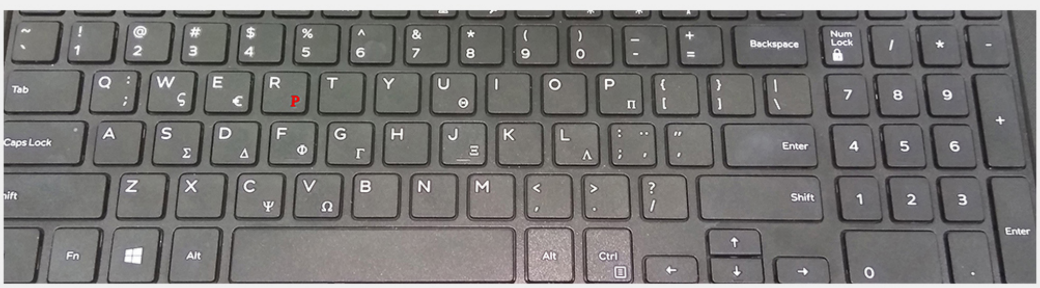
To familiarize yourself with keyboard input of letters, in the next screens you will see a picture of a keyboard identical to the one of the computer you are using.

Your task is to use the keyboard to type the letter colored with red.

ATTENTION: The keyboard, by default, types Greek characters. You do not need to take any additional actions. In addition, it does not matter if you type in small or capital letters.

Continue >>

Screens 5-21
(note: this screen was repeated as many times as the consonant letters of the Greek alphabet i.e., 17 times)

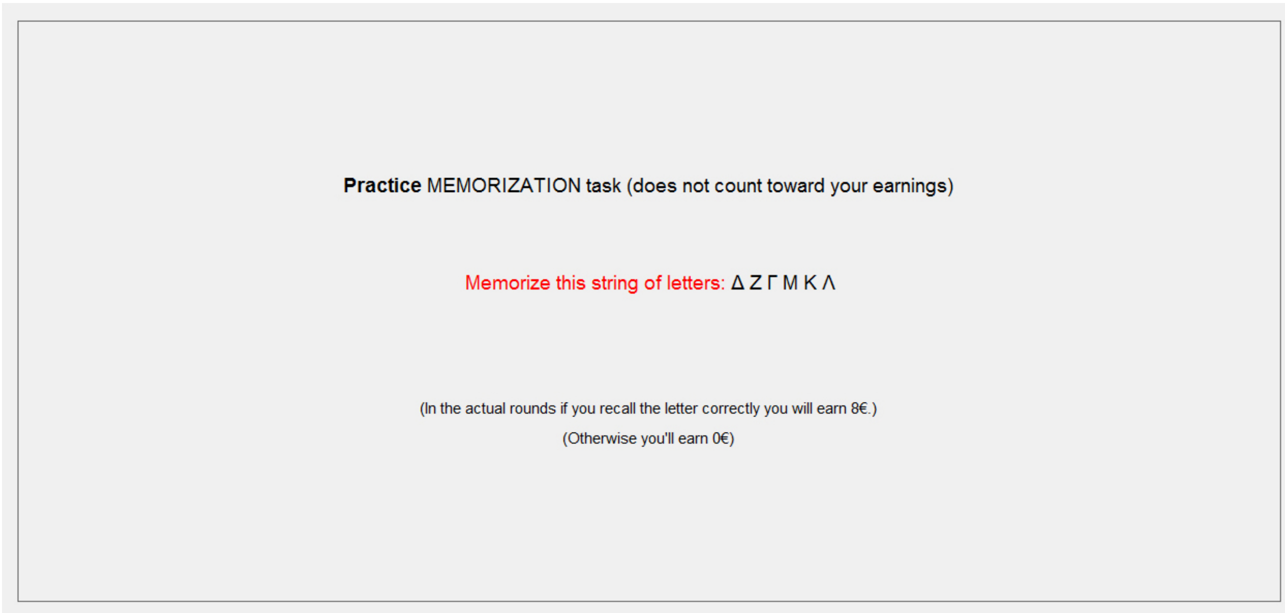


Use the keyboard to type the letter shown in red in the picture above.

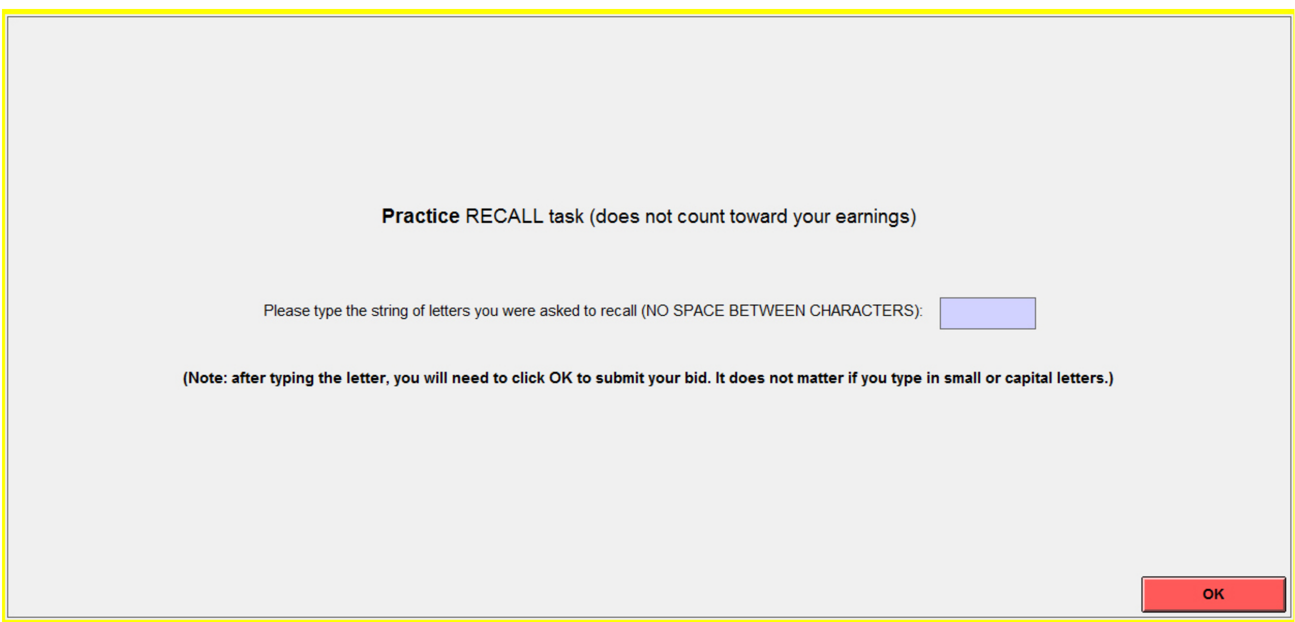
Continue >>

Screen 22

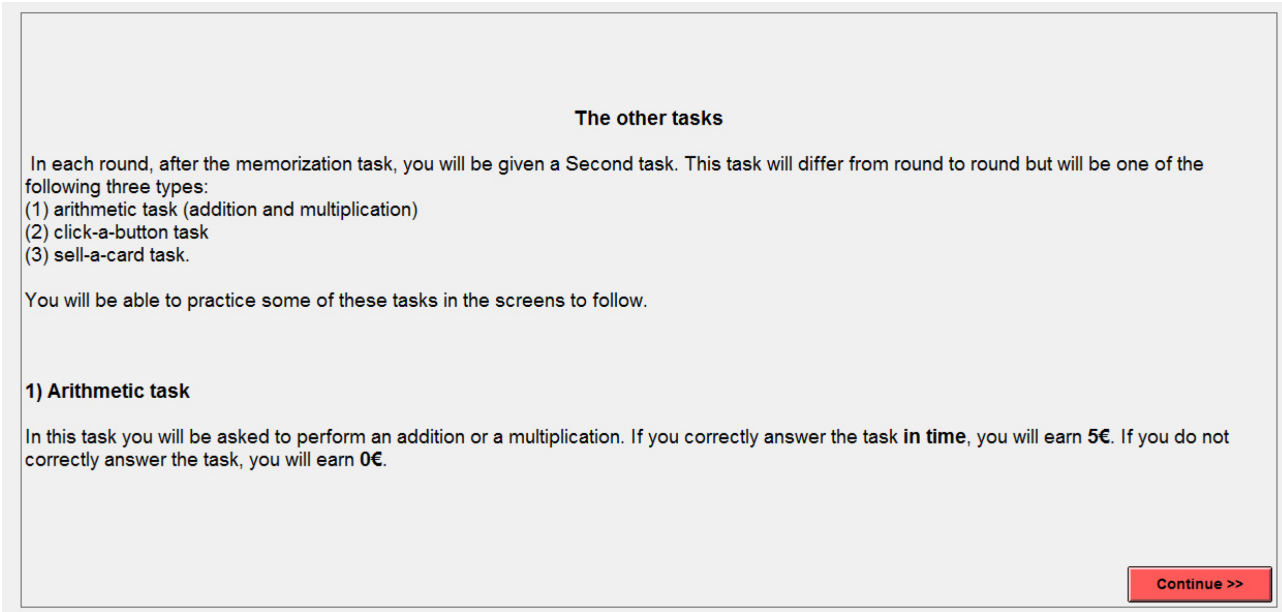
(note: subjects in the low cognitive load treatment were shown one letter instead of a string of letters)



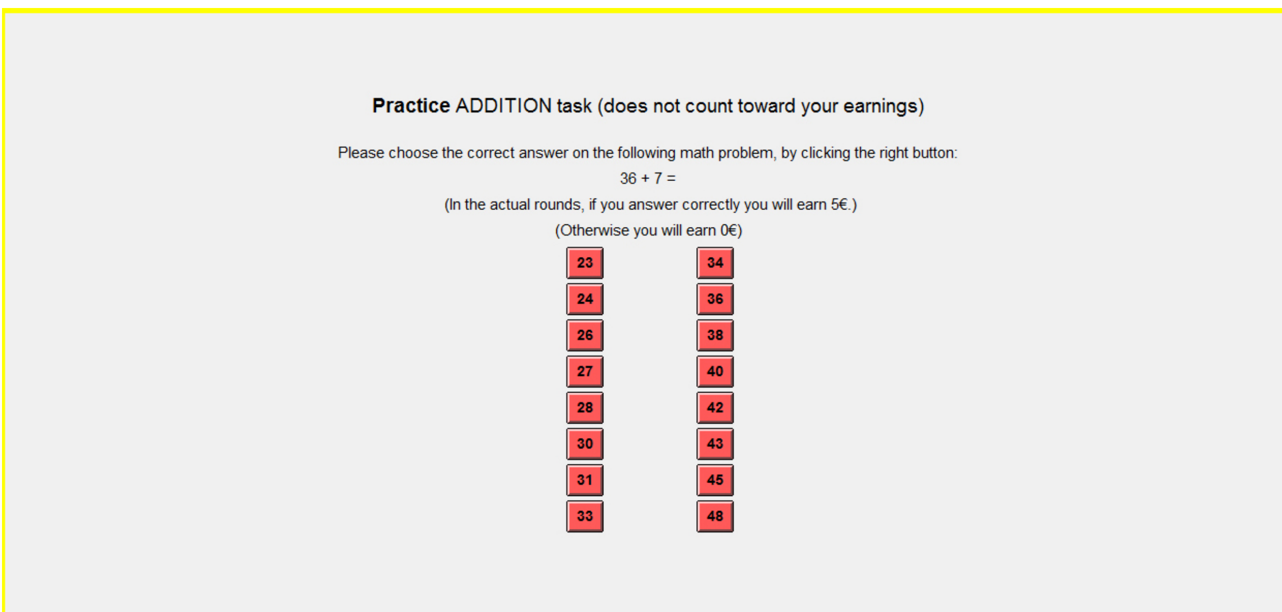
Screen 23



Screen 24



Screen 25



Screen 26

Practice MULTIPLICATION task (does not count toward your earnings)

Please choose the correct answer on the following math problem, by clicking the right button:

14 * 20 =

(In the actual rounds, if you answer correctly you will earn 5€.)

(Otherwise you will earn 0€)

270	279
271	280
272	284
273	287
275	289
276	291
277	296
278	298

Screen 27

2) Click-a-button task

In this task you will be asked to click a button. If you click the button **on time**, you will earn **5€**. If you do not click the button on time you will earn **0 €**.

Continue >>

Screen 28

3) Sell-a-card

In this task you will be asked to sell a card which will appear on your computer screen. Selling the card can earn you money. The rules of selling the card will be written in detail on your screen when you reach the respective stage. Please pay attention because you will need to read instructions about how to sell the card while performing the memorization and recall task.

Continue >>

Screen 29

Summary

In this study there will be 8 rounds. In every round you will perform two tasks, therefore in total you will do 16 tasks as follows:

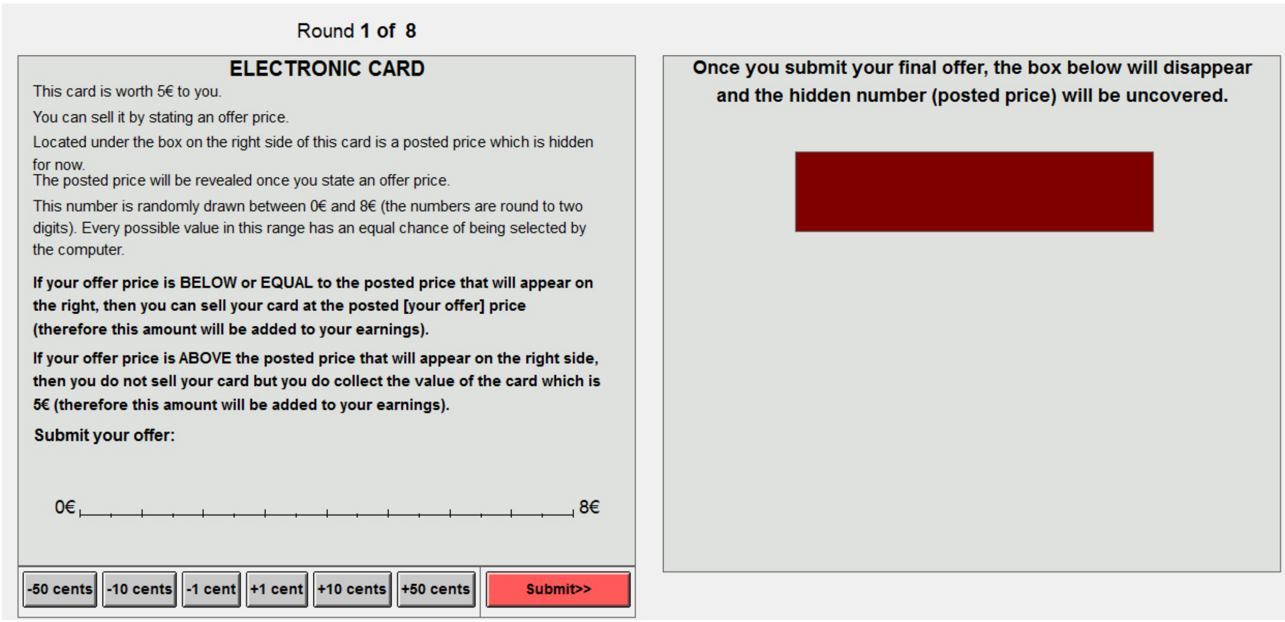
1) **8 times** the MEMORIZATION & RECALL task
2) **2 times** the SELL-A-CARD task
3) **2 times** the ADDITION task
4) **2 times** the MULTIPLICATION task
5) **2 times** the CLICK-A-BUTTON task

Your **additional** earnings *if* this task is randomly selected as binding, will be as follows
1) MEMORIZATION & RECALL task: **8€**
2) SELL-A-CARD task: **your potential earnings will be described once you reach that stage**
3) ADDITION task: **5€**
4) MULTIPLICATION task: **5€**
5) CLICK-A-BUTTON task: **5€**

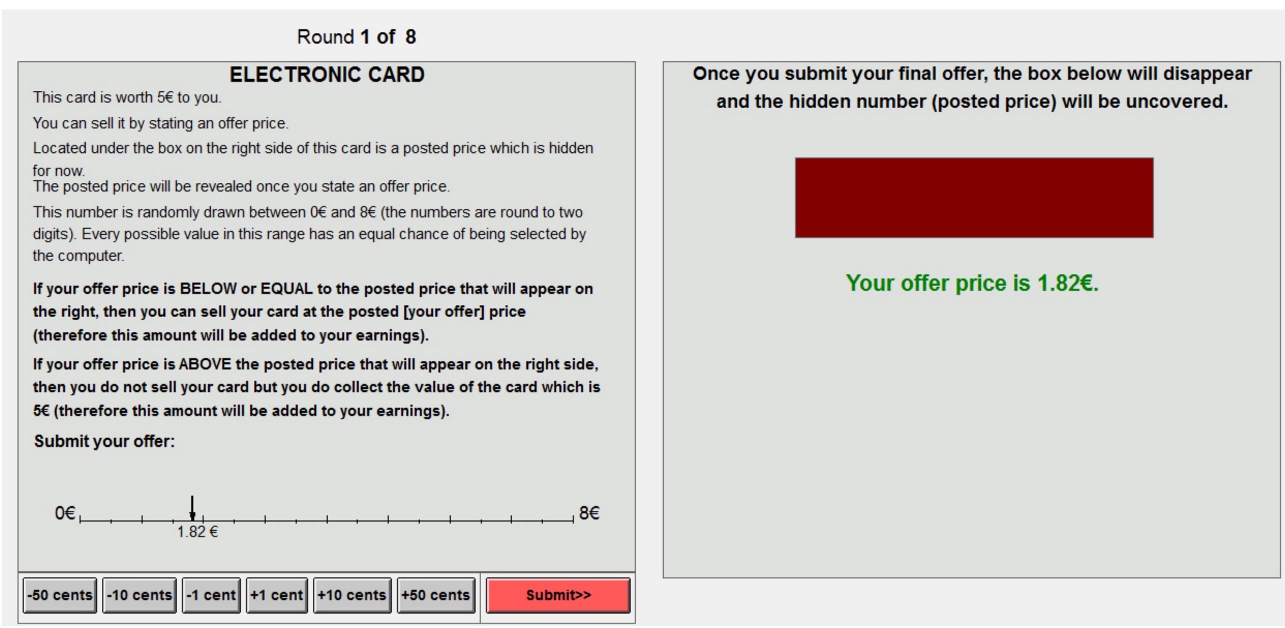
Continue >>

Sample screen shots of the Valuation task

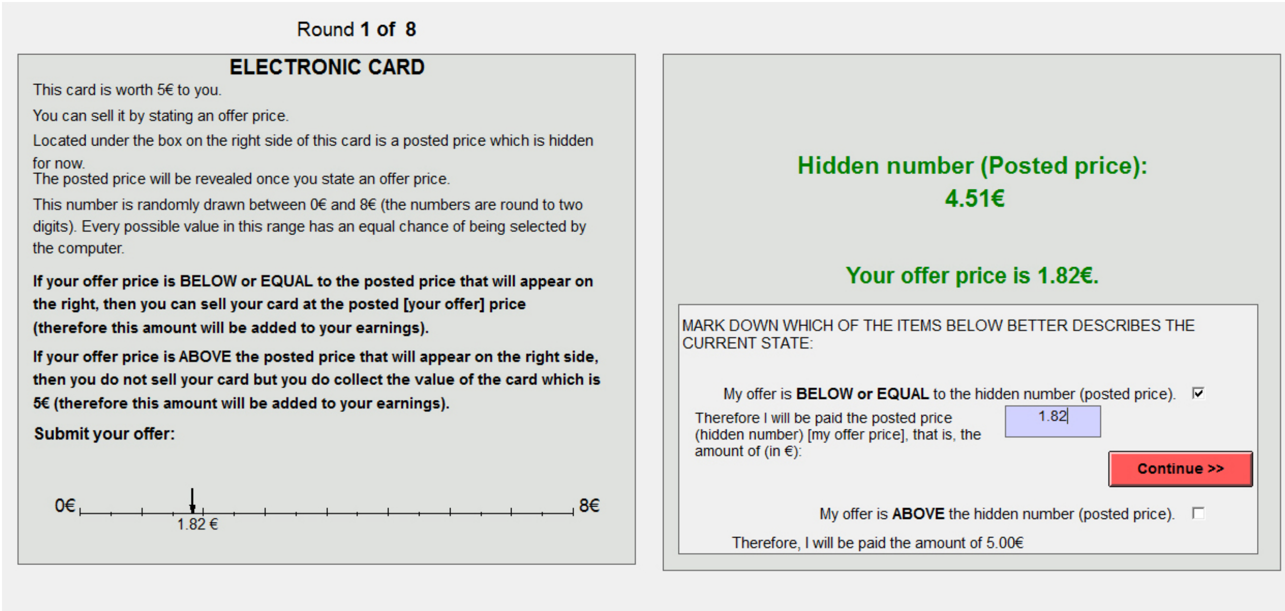
(note: text in brackets indicates differences in the BDM and First Price auction treatments)



(Stating an offer)



(Finalized offer and posted price revealed)



Additional results: Difficulty of the memorization task

In order to explore the differences in recall success while taking into account the influence of other control variables, we estimate a logit model for the success/failure of recalling the memorized letter/string, with clustered standard errors to take into account the multiple responses given by the same subject and to allow for correlation between responses.[3]

Table A1 exhibits the results from a logit regression where success/failure is the dependent variable (model (1)) and a model that interacts the HCL treatment dummy with the tasks to explore differential effects. Model (2) supports the existence of a differential effect of cognitive load on recall success. Because interaction terms complicate the interpretation of the interacted variables in non-linear models, Figure A1 presents graphically the marginal effects for cognitive load and the task dummies. We focus on Figure A1b which shows how difficult memorizing a string was. Subjects exhibit a statistically significant lower probability of recalling the string correctly in the six letter string condition (HCL treatment) than in the one letter condition (LCL treatment). There is some variation between tasks: successfully recalling the string was less likely after the multiplication task than after the click-a-button task.[4]

[3]The robust estimator of variance relaxes the independence assumption and requires only that the observations be independent across the clusters. It involves a slight modiffication of the robust (or sandwich) estimator of variance which requires independence across all observations (StataCorp, 2013, pp. 312).

[4]Figure A1a shows that the effect of the various decision tasks do not differ with respect to probability of correct recall when under LCL with the exception of the Click-a-button task where subjects have a 3.6% higher probability of recalling the letter correctly. When under HCL, subjects are more likely to recall the string correctly after the Click-a-button task (compared to the multiplication task) than after the addition task (19.6% more likely) and after the valuation task (11.2% more likely). Overall, Figure A1a indicates that the probability of correctly recalling the string is higher after less demanding decision tasks.

Table A1: Logit regressions of recall success (clustered standard errors)

	Without interactions		With interaction terms	
	(1)		(2)	
Constant	3.539*	(1.942)	3.749*	(1.978)
Task: Addition	0.891***	(0.169)	0.720	(0.533)
Task: Click-a-button	1.045***	(0.164)	1.836***	(0.664)
Task: Valuation	0.417***	(0.156)	-0.306	(0.365)
HCL treatment	-4.152***	(0.237)	-4.414***	(0.366)
Addition × HCL treatment	-	-	0.222	(0.565)
Click-a-button × HCL treatment	-	-	-0.787	(0.690)
Valuation × HCL treatment	-	-	0.881**	(0.402)
Raven score	0.007	(0.021)	0.007	(0.021)
Period	0.367***	(0.097)	0.367***	(0.097)
Demographics	Yes		Yes	
N	2152		2152	
Log-likelihood	-834.202		-830.342	
AIC	1690.405		1688.684	
BIC	1752.820		1768.123	

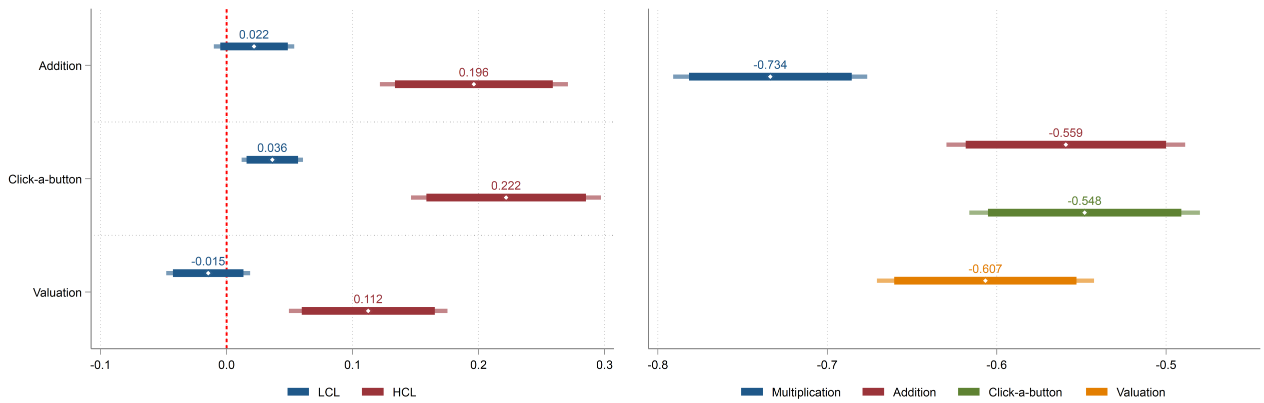
Notes: Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01. Base category is the multiplication task.

The Period variable is positive and statistically significant, indicating that subjects perform better as the experiment progresses.

Figure A1: Marginal effects from Logit model with interaction terms in Table A1 (with 95% and 90% CI)

(a) Marginal effect of tasks conditional on cognitive load

(b) Marginal effect of cognitive load conditional on eash task



Additional results: Manipulation checks

Table A2 shows a logit regression of success/failure at the decision tasks, pooling data together from the three tasks (standard errors are clustered at the individual level) in order to econometrically control for the influence of observable characteristics and to explore the joint influence of the treatment variable and decision tasks. Table A2 shows the results of two specifications, with and without interaction terms. Information criteria favor the model without interaction terms (as well as the fact that none of the interaction terms is statistically significant).

Table A2: Logit regressions of success/failure in the decision tasks

	Without interaction terms (1)		With interaction terms (2)	
Constant	1.450	(1.505)	1.665	(1.631)
Task: Addition	-2.610***	(0.432)	-2.799***	(0.745)
Task: Multiplication	-4.652***	(0.428)	-4.897***	(0.734)
HCL treatment	-0.305**	(0.147)	-0.664	(0.873)
Task: Addition×HCL treatment	-	-	0.298	(0.914)
Task: Multiplication×HCL treatment	-	-	0.402	(0.901)
Raven's test score	0.056***	(0.016)	0.056***	(0.016)
Period	0.380***	(0.145)	0.379***	(0.145)
Demographics	Yes		Yes	
N	1614		1614	
Log-likelihood	-596.533		-596.389	
AIC	1213.066		1216.777	
BIC	1266.930		1281.415	

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01. Base category is the Click-a-button task.

Theory and econometrics of risk preferences

We assume a constant relative risk aversion (CRRA) utility function $U(M) = \frac{M^{1-r}}{1-r}$, where M is a monetary payoff and r is the relative risk aversion (RRA) coefficient. Assuming that Expected Utility Theory (EUT) describes subjects' risk preferences, then the expected utility of lottery i can be written as $EU_i = \sum_{j=1,2} p_i(M_j)U(M_j)$ where $p(M_j)$ are the probabilities for each outcome M_j that are induced by the experimenter in the lottery tasks (see Table A3). We assume subjects have some latent preferences over risk which are linked to observed choices via a probabilistic model function of the general form: $\Pr(B) = F\left(\mu\frac{(V_B - V_A)}{D}\right)$. $\Pr(B)$ is the probability of choosing lottery B (the right hand side lottery in the risk preference tasks), μ is a structural 'noise parameter' (sometimes called a scale or precision parameter) used to allow some errors from the perspective of the deterministic model and V_A, V_B are the decision-theoretic representations of values associated with lotteries A and B i.e., $V_j = EU_j$ for $j = A, B$. $F : R \rightarrow [0, 1]$ is an increasing function with $F(0) = 0.5$ and $F(x) = 1 - F(-x)$, which is to say that this function takes any argument between $\pm\infty$ and transforms it to a number between 0 and 1 i.e., a probability. The F function comes into two flavors in the respective literature: the cumulative standard normal distribution function Φ (the probit link) and the standard logistic distribution function Λ with $\Lambda(\zeta) = 1/(1 + e^{-\zeta})$ (the logit link). D adjusts the scale parameter in heteroskedastic models.

One class of models can be derived when we restrict $D = 1$. This is a class of homoskedastic latent index models also known as Fechnerian or Strong utility models (see Drichoutis and Lusk, 2014). The model with the logit link is equivalent to $\Pr(B) = \Lambda(\mu(V_B - V_A)) = \frac{exp(\mu V_B)}{exp(\mu V_A)+exp(\mu V_B)}$. Another type of the homoskedastic class of models, called Luce or Strict utility models, uses the difference between the logarithm of values: $\Pr(B) = \Lambda(\mu(ln[V_B] - ln[V_A]))$ which is equivalent to $\Pr(B) = \frac{(V_B)^\mu}{(V_A)^\mu+(V_B)^\mu}$.

Table A3: The lottery tasks

Holt and Laury (2002) task								Payoff-varying task							
Lottery A				Lottery B				Lottery A				Lottery B			
p	€	p	€	p	€	p	€	p	€	p	€	p	€	p	€
0.1	2	0.9	1.6	0.1	3.85	0.9	0.1	0.5	1	0.5	1	0.5	1.2	0.5	0.2
0.2	2	0.8	1.6	0.2	3.85	0.8	0.1	0.5	1.2	0.5	1	0.5	1.5	0.5	0.2
0.3	2	0.7	1.6	0.3	3.85	0.7	0.1	0.5	1.4	0.5	1	0.5	1.8	0.5	0.2
0.4	2	0.6	1.6	0.4	3.85	0.6	0.1	0.5	1.6	0.5	1	0.5	2.2	0.5	0.2
0.5	2	0.5	1.6	0.5	3.85	0.5	0.1	0.5	1.8	0.5	1	0.5	2.9	0.5	0.2
0.6	2	0.4	1.6	0.6	3.85	0.4	0.1	0.5	2.0	0.5	1	0.5	3.5	0.5	0.2
0.7	2	0.3	1.6	0.7	3.85	0.3	0.1	0.5	2.2	0.5	1	0.5	4.6	0.5	0.2
0.8	2	0.2	1.6	0.8	3.85	0.2	0.1	0.5	2.4	0.5	1	0.5	6.8	0.5	0.2
0.9	2	0.1	1.6	0.9	3.85	0.1	0.1	0.5	2.6	0.5	1	0.5	9.2	0.5	0.2
1	2	0	1.6	1	3.85	0	0.1	0.5	2.8	0.5	1	0.5	15	0.5	0.2

A second class of models, the heteroskedastic class, can be derived when $D \neq 1$. Wilcox (2008, 2011) proposed a 'contextual utility' error specification which adjusts the scale parameter by $D = V_{max} - V_{min}$ to account for the range of possible outcome utilities. D is defined as the maximum utility V_{max} over all prizes in a lottery pair, minus the minimum utility V_{min} over all prizes in the same lottery pair. It changes from lottery pair to lottery pair, and thus it is said to be contextual. Contextual utility maintains that the error specification is mediated by the range of possible outcome utilities in a pair, so that $\Pr(B) = F\left(\mu \frac{(V_B - V_A)}{V_{max} - V_{min}}\right)$.

Another heteroskedastic model which has received some attention in economics lately (Hey et al., 2010; Wilcox, 2015) is prescribed by Decision Field Theory (DFT) (Busemeyer and Townsend, 1992, 1993). DFT allows the decision maker's attention to switch from one event to another across choice pairs. This variability on focus on events is caused by a random difference which Busemeyer and Townsend (1993) name a *valence difference*. The variance of this valence difference in the case of lotteries with just two outcomes is given by $D^2 = w(p_1)(V_{A1} - V_{B1})^2 + (1 - w(p_1))(V_{A2} - V_{B2})^2 - (V_A - V_B)^2$ where V_{A1}, V_{A2}, V_{B1} and V_{B2} are the representations of values associated with the first and second outcome of lottery A and B, respectively; V_A and V_B are the representations of values associated with lottery A and B, respectively. Note that when lotteries are certainties, such as in the last row of the HL task, then $D = 0$ and $\Pr(B) = 1$, that is the subject always chooses the dominating lottery. This also implies that the last row of the HL task must be excluded from estimation under DFT.

After defining the decision theoretical models and error specifications, the log-likelihood function can then be written as: $\ln L(w_i) = \sum_{i=1}^{N} \left[(\ln Z | w_i = 1) + (\ln(1 - Z) | w_i = -1)\right]$, where $Z = Pr_j$ and j indexes the different error models (j =FP, FL, STRICT, CP, CL, DFTP, DFTL).[5] $w_i = 1$ denotes the choice of lottery B and $w_i = -1$ denotes the choice of the A lottery in the risk preference task i. Subjects were allowed to express indifference between choices and were told that if that choice was selected to be played out, the computer would randomly choose one of the two options for them and that both choices had equal chances of being selected. The

[5]FP and FL stand for the Fechner error with a probit and a logit link, respectively. CP and CL stand for contextual utility with a probit and a logit link, respectively. DFTP and DFTL stand for Decision Field theory with a probit and a logit link respectively. STRICT stands for Luce error or Strict utility.

likelihood function for indifferent choices is constructed such that it implies a 50/50 mixture of the likelihood of choosing either lottery so that: $\ln L(w_i) = \sum_{i=1}^{N} \left[(\ln Z|w_i = 1) + (\ln(1-Z)|w_i = -1) + (\frac{1}{2}\ln Z + \frac{1}{2}\ln(1-Z)|w_i = 0)\right]$ where $w_i = 0$ denotes the choice of indifference. We can then use standard numerical methods to maximize the log-likelihood function.

In order to select between the competing stochastic models, we used Akaike's and Bayesian information criteria (AIC and BIC). AIC and BIC do not reveal how well a model fits the data in an absolute sense, i.e., there is no null hypothesis being tested. Nevertheless, these measures offer relative comparisons between models on the basis of information lost from using a model to represent the (unknown) true model.[6]

Table A4: Akaike and Bayesian Information criteria by error story

		AIC	AIC corrected	BIC
Fechner	Logit	3886.875	3887.125	4025.671
	Probit	3928.770	3929.021	4067.567
Contextual utility	Logit	3695.817	3696.068	3834.613
	Probit	3715.654	3715.905	3854.451
Decision Field Theory	Logit	3698.825	3698.950	3792.689
	Probit	3756.337	3756.462	3850.202
Strict utility		3898.062	3898.313	4036.859

As shown in Table A4, the Contextual utility model with a logit link (CL) is the preferred model according to AIC and Decision Field Theory with a logit link (DFTL) is the preferred model according to BIC. Because DFTL produces very close AIC values to CL, we use DFTL to estimate the parameters of the model. However, the predicted values for r of the CL model are extremely close to those of the DFTL model (pairwise correlation coefficient is 0.996) which renders selection of one model over the other a trivial task. Results from estimating the DFTL model are shown in Table A5. We use these estimates to predict the coefficient of relative risk aversion for 203 subjects and predict out of sample the r values for the rest of the 66 subjects.

[6] Drichoutis and Lusk (2016) have shown that AIC and BIC are in agreement in terms of model selection with more complex selection criteria such as Vuong's (1989) test, Clarke's (2003) test or the out-of-sample log likelihood (OSLLF) criterion (Norwood et al., 2004).

Table A5: Estimates for r and μ given the Decision Field theory stochastic assumption

	r		μ	
Constant	0.713	(0.546)	2.935***	(0.176)
Male	-0.134*	(0.077)		
Accepted payment	0.070	(0.112)		
Age	-0.014	(0.021)		
Household size	0.028	(0.037)		
Income group:				
Income_2	-0.012	(0.182)		
Income_3	0.062	(0.148)		
Income_4	0.116	(0.146)		
Income_5	0.080	(0.139)		
Income_6	0.060	(0.144)		
Income_7	-0.145	(0.299)		
N	3857			
Log-likelihood	-1834.412			

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01.

Joint estimation of r and λ

We can write the conditional log-likelihood from the valuation task as (see the 'optimal offers under risk section' in the paper):

$$\ln L^m(\lambda; y_i) = \sum_i \ln \frac{y_i e^{\lambda EU^m[\pi|b_j]}}{\sum_n^{k=1} e^{\lambda EU^m[\pi|b_k]}} \tag{7}$$

and the conditional log-likelihood from the risk preferences task as:

$$\ln L^{RA}(w_i) = \sum_{i=1}^{N} \left[(\ln Z|w_i = 1) + (\ln(1-Z)|w_i = -1) + (\frac{1}{2}\ln Z + \frac{1}{2}\ln(1-Z)|w_i = 0) \right] \tag{8}$$

The joint likelihood of the risk preferences choices and the valuation task can then be written as:

$$\ln L(\lambda, r, \mu; y, w) = \ln L^{RA} + \ln L^m \tag{9}$$

Sample size calculations

Our per treatment sample size was decided based on sample size calculations and served as a stopping rule for this experiment when we achieved the minimum necessary per treatment sample. Assuming $\alpha = 0.05$ (Type I error) and $\beta = 0.20$ (Type II error), the per group (treatment) minimum sample size required to compare two means μ_0 and μ_1, with common variance of σ^2 in order to achieve a power of at least $1 - \beta$ is given by (Diggle et al. (2002) pp. 30; Liu and Wu (2005); Kupper and Hafner (1989)):

$$n = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2(1 + (M-1)\rho)}{M(\frac{\mu_0-\mu_1}{\sigma})^2} \quad (10)$$

To take into account the repeated measurement, the formula includes the number of repeated measurements M (in our case it is $M = 2$) as well as a value for the correlation ρ between observations for the same subject. For $\alpha = 0.05$ and $\beta = 0.20$ the values of $z_{1-\alpha/2}$ and $z_{1-\beta}$ are 1.96 and 0.84, respectively. To calculate a minimum sample size, one needs to feed the above formula with values for σ and the minimum meaningful difference $d = \mu_0 - \mu_1$. To specify the necessary parameters to feed the above formula, we looked at the values for σ reported in Table 1 in Bull et al. (2019) which range from 0.9 to 1.15. In addition, we used values for ρ spanning the range from 0.1 to 0.7 (larger values of correlation are unlikely unless all subjects submit (almost) the same bid in Period 2 as in Period 1 or bids in Period 2 vary uniformly among subjects). The minimum effect size we considered was a difference of 50 cents.

Table A6 shows the result of equation 10 for various values of σ, ρ and d. It is evident that the lower the minimum meaningful difference d, the higher the correlation between periods ρ and the higher the standard deviation σ, a larger sample size is needed to detect the desired effect size with 80% power. Our per cell sample size can likely detect a minimum meaningful difference of 50 cents for various values of σ and ρ. We can also detect smaller differences than 50 cents but one would need to restrict the range of assumed values for σ and ρ.

Table A6: Per treatment sample size calculations for different values of σ, ρ and d

		$\sigma = 0.9$	$\sigma = 1$	$\sigma = 1.15$
	$\rho = 0.1$	28	35	46
$d = 0.50$	$\rho = 0.5$	38	47	62
	$\rho = 0.7$	43	53	71
	$\rho = 0.1$	23	29	38
$d = 0.55$	$\rho = 0.5$	32	39	51
	$\rho = 0.7$	36	44	58
	$\rho = 0.1$	19	24	32
$d = 0.60$	$\rho = 0.5$	26	33	43
	$\rho = 0.7$	30	37	49

Additional tables

Table A7: Possible outcomes under different bidding strategies

		b ≤ posted price	b > posted price	max Payoff	min Payoff
	b < 5	[0, 8]	5	*8*	0
BDM	b = 5	[5, 8]	5	8	*5*
	b > 5	(5, 8]	5	*8*	5
	b < 6.5	[0, 6.5)	5	*6.5*	0
FPA	b = 6.5	6.5	5	6.5	*5*
	b > 6.5	(6.5, 8]	5	*8*	5

Table A8: Comparing success rate in the recall task with other studies

		HCL	LCL
	Deck and Jahedi (2015)	55.90%	71.60%
Multiplication	Drichoutis and Nayga (2017)	40.58%	53.68%
	This study	44.85%	51.50%
	Deck and Jahedi (2015)	96.90%	97.80%
Addition	Drichoutis and Nayga (2017)	87.21%	91.70%
	This study	84.92%	89.10%
Click-a-button	Drichoutis and Nayga (2017)	99.30%	99.42%
	This study	98.53%	99.25%

Notes: HCL (LCL) stands for the high (low) cognitive load treatment.

Table A9: Maximum likelihood estimates of logit choice error parameter λ for first price auction misconception and mixture models (risk neutrality assumed away)

	Joint estimation with r	
	FP-GFM model	Mixture model
	(1)	(2)
λ		
Constant	0.302*	1.138**
	(0.181)	(0.512)
HCL treatment	-0.253	-0.851
	(0.252)	(0.640)
π_{GFM}		
Constant		≈0.000
		(<0.001)
HCL treatment		≈0.000
		(<0.001)
r		
Constant	0.872	0.893
	(0.858)	(0.853)
Male	-0.117	-0.120
	(0.114)	(0.114)

Accepted payment	0.107	0.112
	(0.189)	(0.194)
Age	-0.016	-0.017
	(0.030)	(0.030)
Household size	0.025	0.025
	(0.050)	(0.049)
Income		
$Income_2$	0.046	0.062
	(0.188)	(0.188)
$Income_3$	-0.017	-0.019
	(0.181)	(0.178)
$Income_4$	-0.052	-0.045
	(0.164)	(0.162)
$Income_5$	-0.046	-0.047
	(0.151)	(0.148)
$Income_6$	-0.120	-0.120
	(0.153)	(0.150)
$Income_7$	-0.074	-0.068
	(0.257)	(0.258)
μ	2.948***	2.944***
	(0.247)	(0.247)
N [Subjects]	2205 [105]	2205 [105]
Log-likelihood	-1631.639	-1724.777
AIC	3289.278	3481.555
BIC	3363.358	3572.730

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01.

Table A10: Maximum likelihood estimates of logit choice error parameter λ for optimal offers, first price auction misconception and mixture models (sample restricted to those with good stated understanding of instructions)

		Risk neutrality		Risk aversion	
	Optimal model	FPA-GFM model	Mixture model	FPA-GFM model	Mixture model
	(1)	(2)	(3)	(4)	(5)
λ					
Constant	1.361***	0.326***	1.253*	0.470**	1.304***
	(0.441)	(0.094)	(0.641)	(0.194)	(0.461)
HCL treatment	-1.038*	-0.199	-1.104	-0.379	-0.992*
	(0.595)	(0.144)	(0.704)	(0.255)	(0.578)
π_{GFM}					
Constant			0.097		0.155
			(0.257)		(0.256)
HCL treatment			0.734		-0.092
			(1.212)		(0.569)
N	222	222	222	222	222
Log-likelihood	-820.949	-821.469	-819.848	-825.211	-820.652
AIC	1645.899	1646.938	1647.696	1654.422	1649.304
BIC	1652.704	1653.743	1661.307	1661.227	1662.915

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01.

Table A11: Maximum likelihood estimates of logit choice error parameter λ for optimal offers, first price auction misconception and mixture models (sample restricted to those with good stated understanding of instructions and payoffs)

		Risk neutrality		Risk aversion	
	Optimal model (1)	FPA-GFM model (2)	Mixture model (3)	FPA-GFM model (4)	Mixture model (5)
λ					
Constant	1.553***	0.361***	1.504***	0.574***	1.493***
	(0.429)	(0.089)	(0.533)	(0.193)	(0.482)
HCL treatment	-0.864	-0.123	-1.102*	-0.305	-0.882
	(0.621)	(0.154)	(0.620)	(0.287)	(0.622)
π_{GFM}					
Constant			0.052		0.172
			(0.191)		(0.260)
HCL treatment			0.534		0.171
			(0.331)		(0.396)
N	208	208	208	208	208
Log-likelihood	-764.013	-764.225	-761.925	-767.770	-763.067
AIC	1532.027	1532.451	1531.850	1539.540	1534.133
BIC	1538.702	1539.126	1545.200	1546.215	1547.484

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05 *** p<0.01.

Additional figures

Figure A2: Probability of FPA-GFM for various relative risk aversion coefficients using Cason and Plott's (2014) data and a mixture model specification

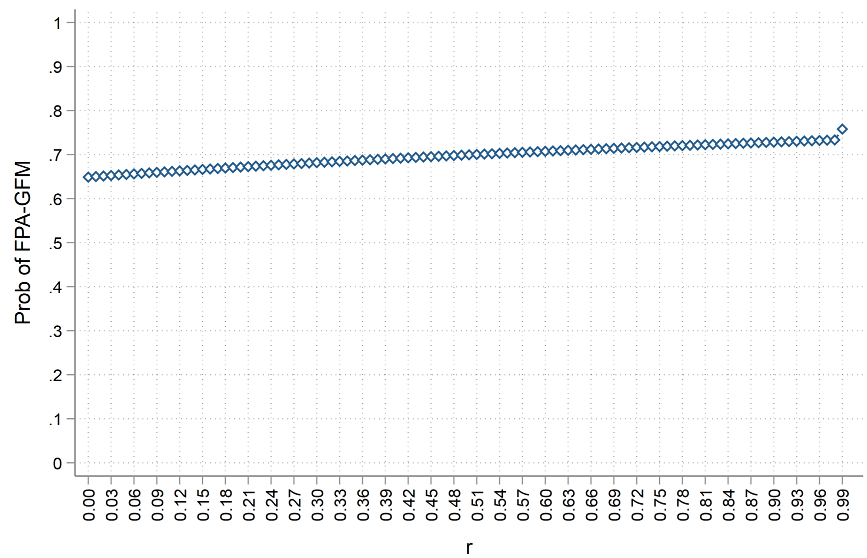
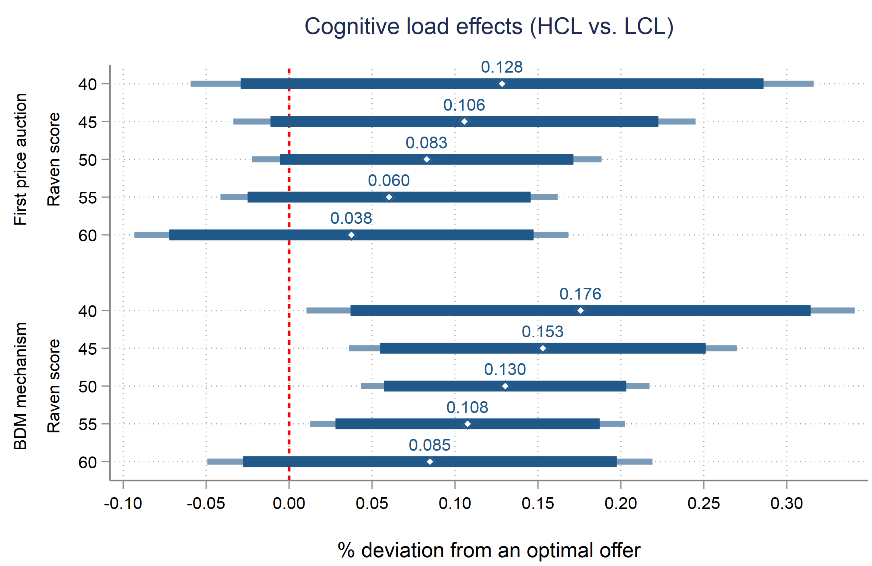
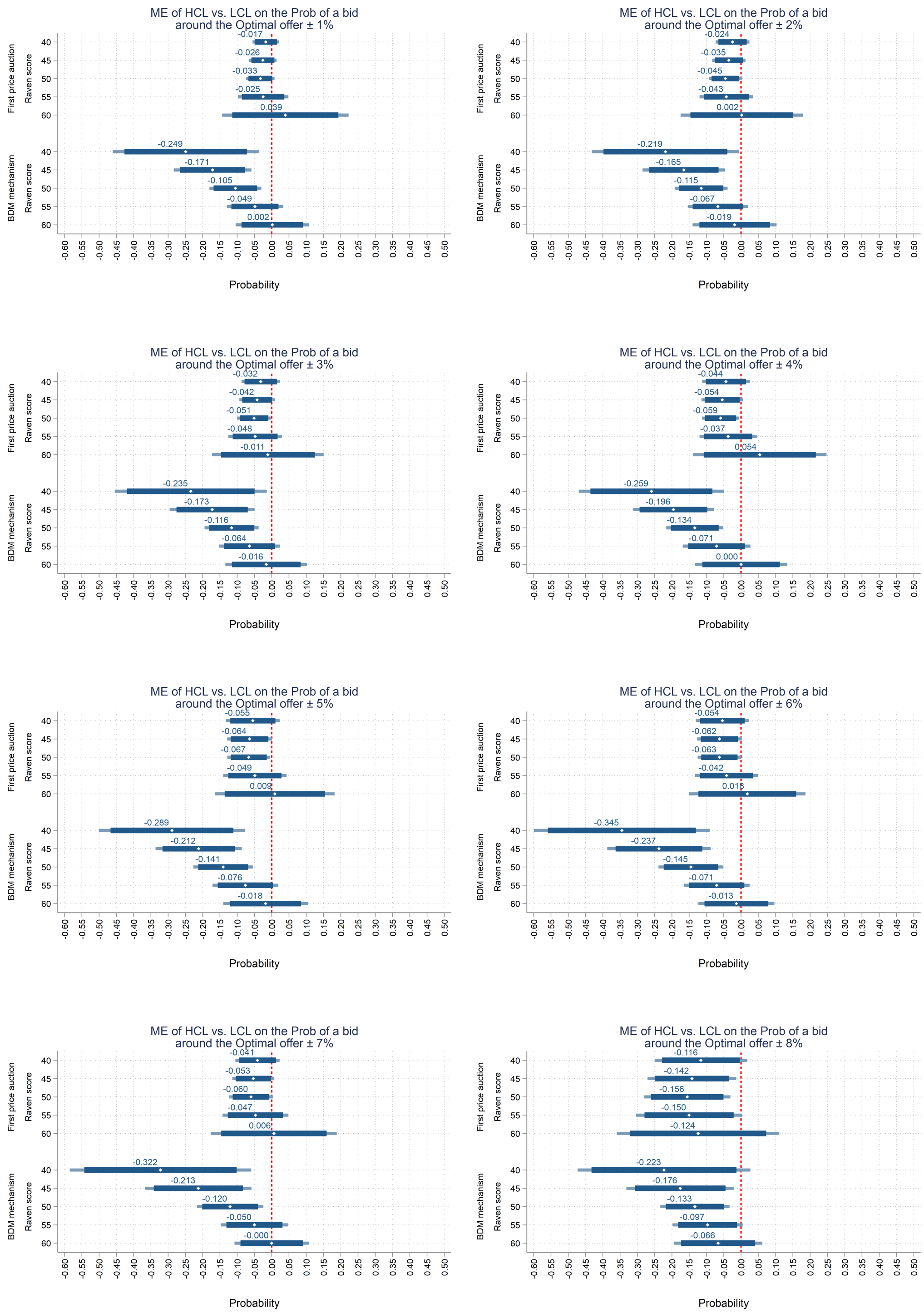


Figure A3: Marginal effects of HCL treatment from RE regression of bid deviations by Valuation mechanism and Raven score (with 95% and 90% confidence intervals; Risk aversion taken into account)



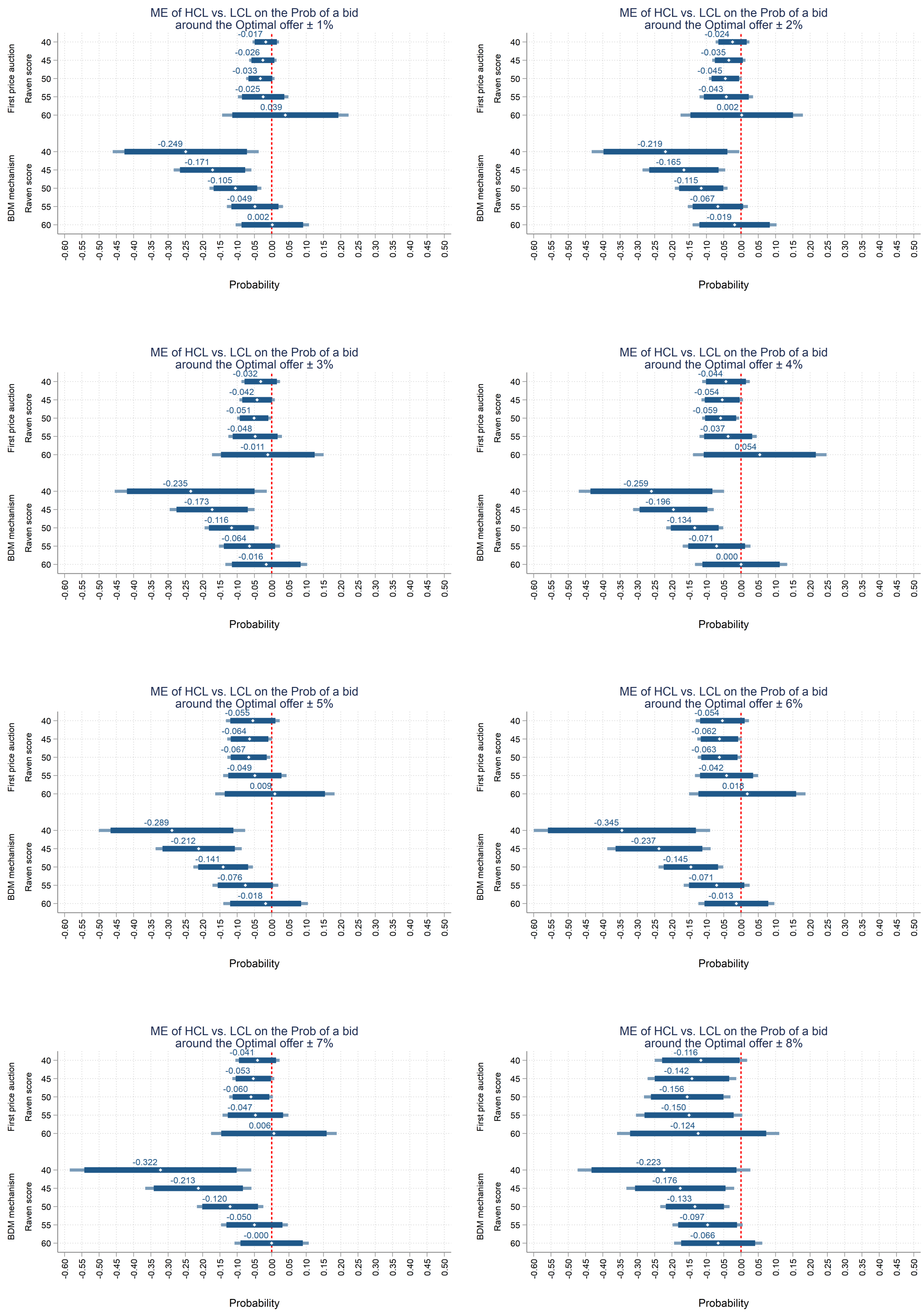
Notes: ME stands for Marginal Effect; RE stands for Random Effects; PML stands for Penalized Maximum Likelihood

Figure A4: Marginal effects of HCL treatment by Valuation mechanism and Raven score for various % deviations from optimal offer (based on estimations from a random effects Logit regression)



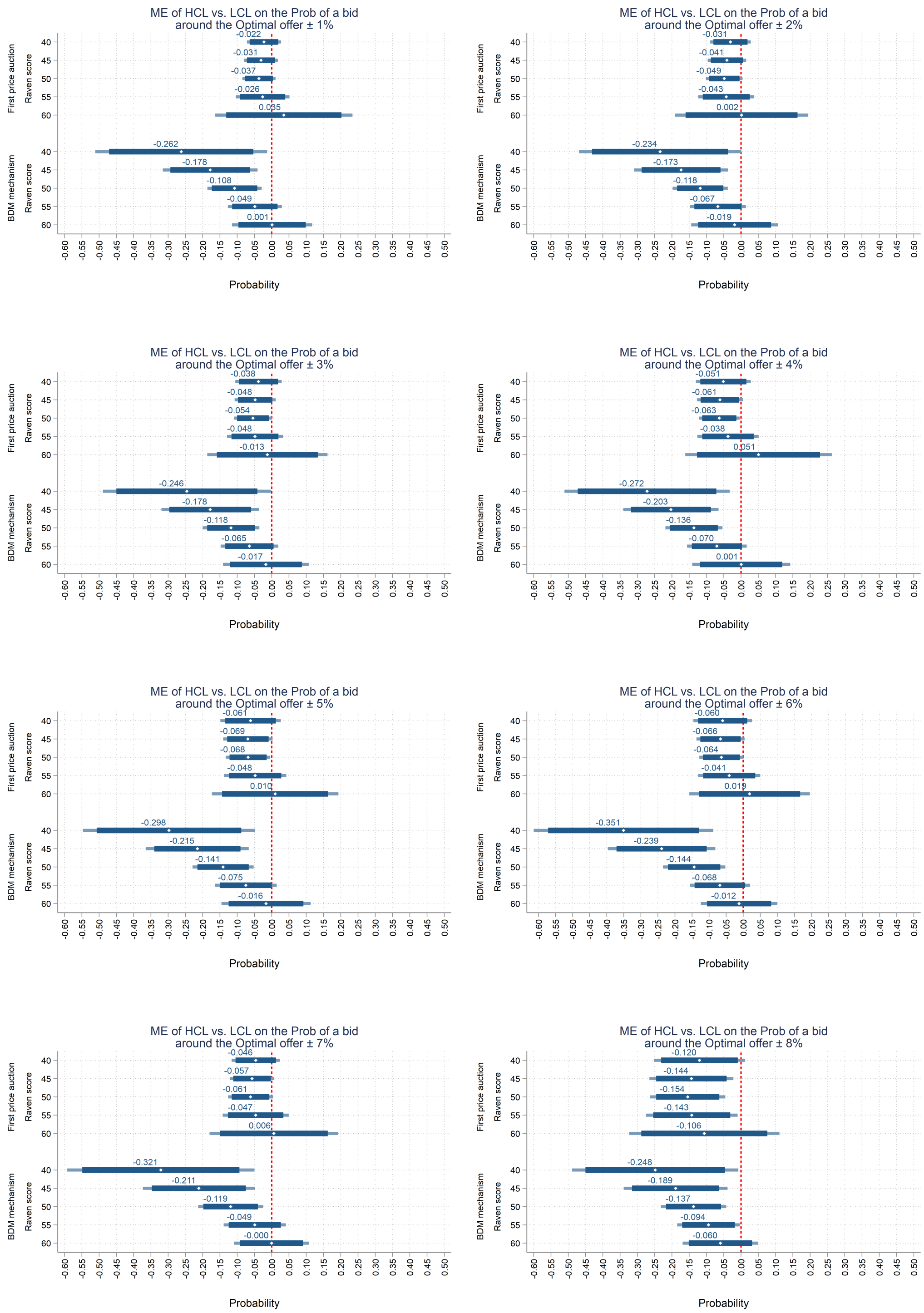
Notes: ME stands for Marginal Effect; HCL (LCL) stands for High (Low) Cognitive Load.

Figure A5: Marginal effects of HCL treatment by Valuation mechanism and Raven score for various % deviations from optimal offer (based on estimations from a random effects Logit regression; Risk aversion taken into account)



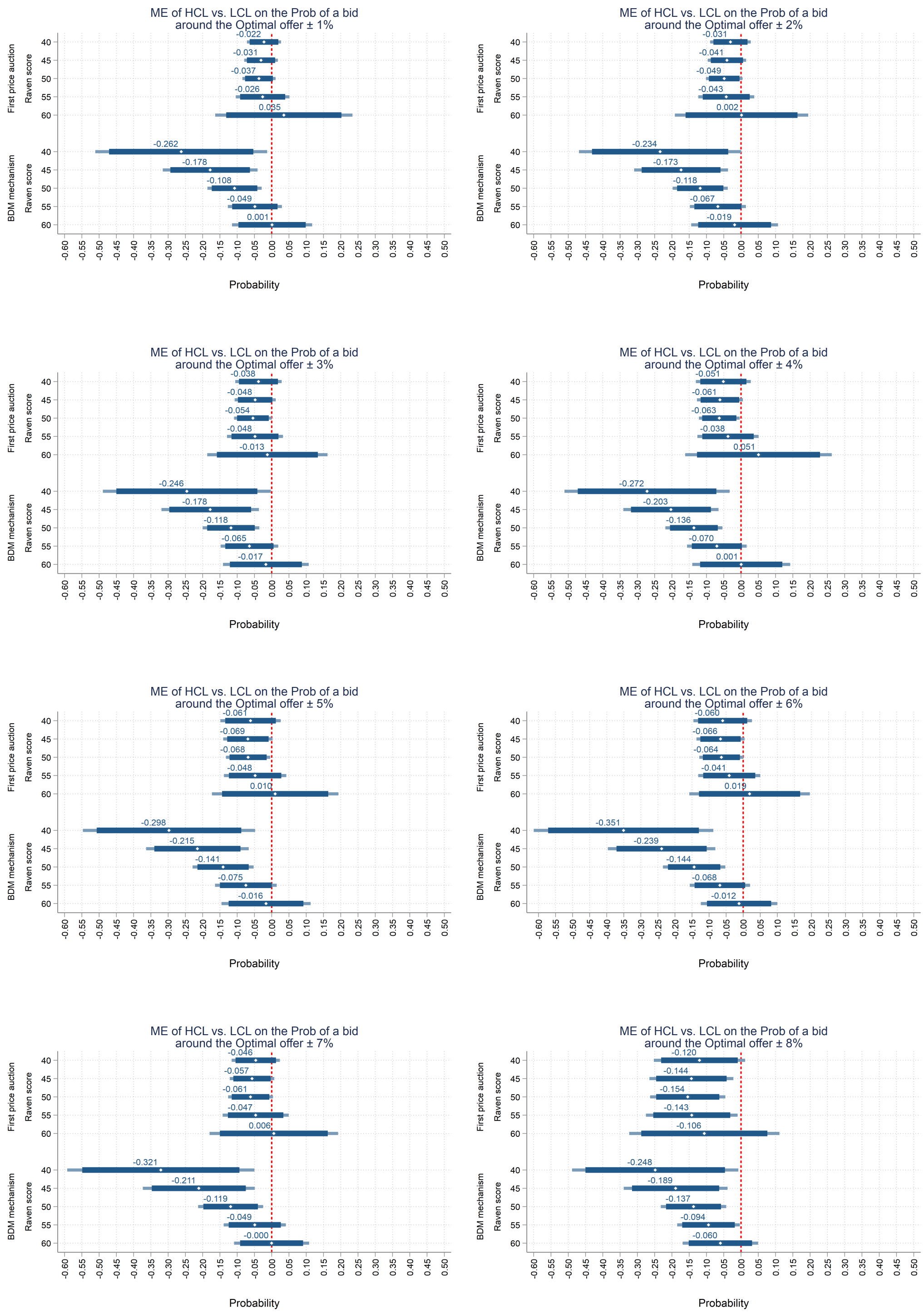
Notes: ME stands for Marginal Effect; HCL (LCL) stands for High (Low) Cognitive Load.

Figure A6: Marginal effects of HCL treatment by Valuation mechanism and Raven score for various % deviations from optimal offer (based on estimations from Penalized Maximum Likelihood Logit regressions)



Notes: ME stands for Marginal Effect; HCL (LCL) stands for High (Low) Cognitive Load.

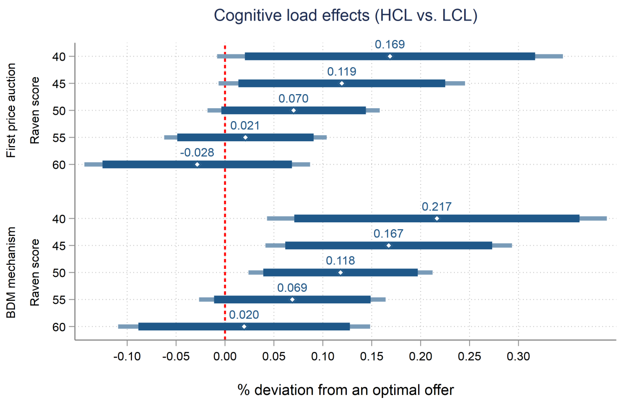
Figure A7: Marginal effects of HCL treatment by Valuation mechanism and Raven score for various % deviations from optimal offer (based on estimations from Penalized Maximum Likelihood Logit regressions; Risk aversion taken into account)



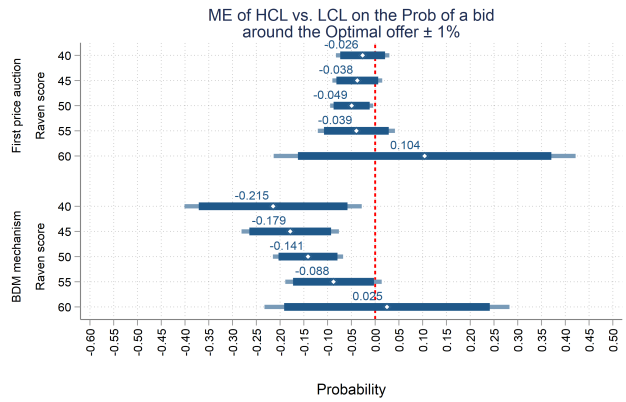
Notes: ME stands for Marginal Effect; HCL (LCL) stands for High (Low) Cognitive Load.

Figure A8: Marginal effects of HCL treatment by Valuation mechanism and Raven score (with 95% and 90% confidence intervals; sample restricted to those with good stated understanding of instructions)

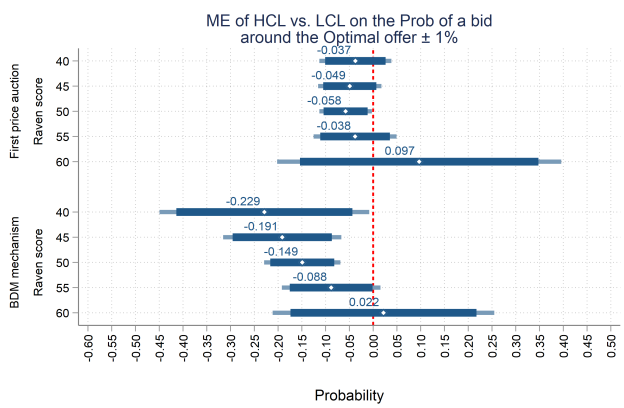
(a) ME from RE regression



(b) ME from RE Logit



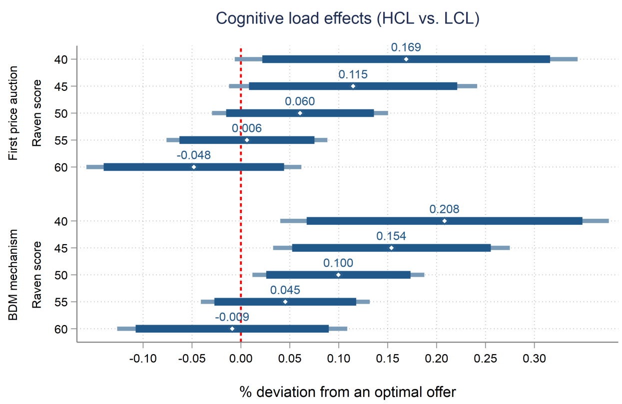
(c) ME from PML Logit



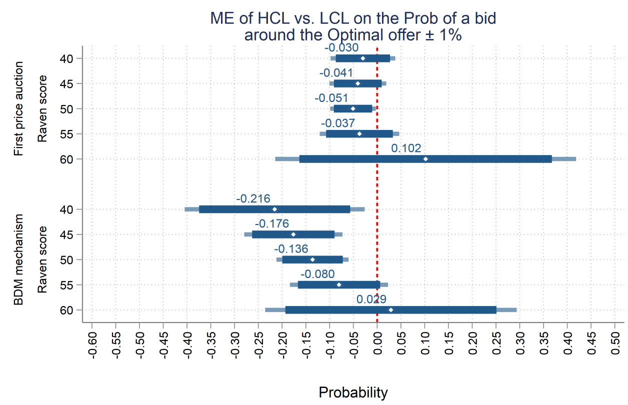
Notes: ME stands for Marginal Effect; RE stands for Random Effects; PML stands for Penalized Maximum Likelihood

Figure A9: Marginal effects of HCL treatment by Valuation mechanism and Raven score (with 95% and 90% confidence intervals; sample restricted to those with good stated understanding of instructions and payoffs)

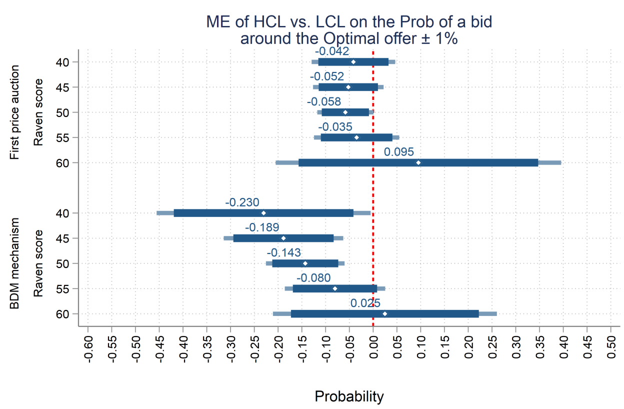
(a) ME from RE regression



(b) ME from RE Logit



(c) ME from PML Logit



Notes: ME stands for Marginal Effect; RE stands for Random Effects; PML stands for Penalized Maximum Likelihood