When does real become consequential in non-hypothetical choice experiments?

Daniel E. Chavez∗ Marco A. Palma† Rodolfo M. Nayga Jr.‡§

March 6, 2019

Abstract: The proneness of stated preference methods to hypothetical bias has increased the popularity of incentivized studies, in particular the use of real choice experiments (RCE). Challenges of RCE include the lack of engagement with the choice task by some subjects, and that some of the product alternatives may not be available in order to incentivize all the choices. This issue brings to question whether the proportion of available products influences the results of the RCE. Would the subjects’ engagement change? Using an induced value choice experiment with a profit maximization optimal strategy for agents, we varied the number of potentially binding alternatives in four treatments. Our results suggest that incentives matter, as the percentage of optimal choices was lowest in the hypothetical treatment. Interestingly, however, we do not find statistically significant differences in the number of optimal choices between the incentivized treatments, regardless of the number of potentially binding alternatives used in our treatments. This suggests that practitioners could conduct incentivized RCE without the need to have all the product alternatives be made available in the study. Furthermore, we explore the interaction of incentives with subjects’ numerical ability and individual reflective state. Both are also shown to influence how incentives impact performance, shedding some light on what individual characteristics to look for when conducting valuation research.

JEL Codes: C91, C18

Keywords: Choice Experiments, Eye Tracking, Hypothetical bias, Induced values

∗Corresponding Author. Department of Marketing and Supply Chain, University of Kentucky, 550 South Limestone, Lexington, KY 40506-0034, USA. Email: daniel.chavez@uky.edu
†Department of Agricultural Economics, Texas A&M University, 2124 TAMU, College Station, TX 77843-2124. Email: mapalma@tamu.edu
‡Agricultural Economics and Agribusiness, 217 Agriculture Building University of Arkansas, University of Arkansas, Fayetteville, AR 72701. Email:rnayga@uark.edu
§We would like to thank the useful comments from seminar participants at Texas A&M and attendees at SAEA.
1 Introduction

Stated preference methods such as choice experiments (CE) are widely used for market and environmental valuation. A well-established literature documents the recurring phenomenon of hypothetical bias in stated preference studies using different mechanisms, including CE (List and Gallet 2001, Murphy et al. 2005, Penn and Hu 2018). Hypothetical bias implies that in the absence of economic incentives there is an overvaluation of products and services (Harrison and Røtstrøm 2008). Because of this issue, researchers are gravitating towards methods that mitigate hypothetical bias, which for market good valuations include incentivized real choice experiments (RCE) (Ding, Grewal, and Liechty 2005, Penn and Hu 2018).

One of the challenges when designing RCE is that some of the product alternatives presented in the choice sets may not be available yet in markets or physically present, which is why incentivizing is not feasible for non-market and environmental valuations. After all, one important research question addressed by CE is how consumers would respond to new products, new features of existing products, or new production technologies. Hence, given that deception is generally not allowed in economic experiments (Colson et al. 2016), researchers using RCE have to truthfully inform respondents that some of the products they are choosing from in the choice sets are not available. This information could modify choice behavior, especially when in RCE –as in other valuation techniques– it is uncertain how much subjects understand what is being asked from them; i.e. do they recognize the form of the game (Cason and Plott 2014). Also, knowing that not all decisions are consequential could change the perceptions of participants and how much effort they are willing to
exert not only to understand the procedure, but also to reveal their true preferences (Yang, Toubia, and de Jong 2018).

Our main research question focuses on whether or not the number of products that will be available to purchase in a RCE affects the subjective perception of subjects on the “realness” of the choices. In other words, we wish to know if subject behavior in RCE is different when the number of product alternatives available for purchase differs. It is an important empirical question whether there is a threshold in the number of alternatives available for purchase that is not salient enough, from a practical perspective, to incentivize subjects in RCE. Are subjects disengaged with the study and unwilling to exert effort to understand and follow the experimental procedures when the incentives provided are not perceived as consequential? In practice, coming up with all the product alternatives in an RCE is a challenge, and some of the alternatives will not be available for purchase. This condition coupled with the fact that CEs are a novel task for most participants in these experiments could play an important role in changing perceptions and effort.

We investigate this question by setting up an induced value choice experiment where the optimal strategy of agents is profit maximization. Because in our design there is a unique known profit-maximizing alternative in each choice set, we can evaluate deviations from the rational profit maximizing strategy resulting from changes in the number of alternatives available for purchase. We set up four cases using a between-subjects design: 1) purely hypothetical where none of the alternatives are available for purchase, 2) 33% of the alternatives are available for purchase, 3) 66% of the alternatives are available for purchase, and 4) all alternatives are available for purchase. We set up the experiment by randomly drawing a subset of the alternatives to make them available for purchase according to each experimental condition. The number of available alternatives is common knowledge to all participants; however, they do not know which alternatives are potentially binding when making their decisions. The full details of the procedure are described in the experimental design section.
To position our research in the valuation literature, and the relevance of hypothetical bias in the setup of our design, we first establish the magnitude of hypothetical bias (i.e., amount of deviation from optimal choices) in an induced value experiment using eye tracking technology. To explain the deviations from optimal rational behavior, we discuss possible sources of hypothetical bias in the review of the existing literature. Specifically, we evaluate the influence of numerical ability, as identifying the optimal profit maximizing alternative requires number reasoning. We also consider the cognitive reflection state and attentiveness on the subjects’ performance as they may selectively ignore or attend information. In order to formally present our research questions, we propose the following hypotheses:

**Hypothesis 1:** *Economic incentives matter in CE.*

The literature has extensively documented the benefit of using incentives in CE. Our design allows measuring the effect of number of available product alternatives on choice behavior in a controlled environment by using an induced value setting. We believe that in line with the extant literature, performance in the choice task will be better in the presence of incentives.

**Hypothesis 2:** *There is a monotonic increase in performance in the choice task with more incentivized decisions.*

As discussed earlier, understanding how the various degrees of availability of the product alternatives in the choice task can have behavioral and value consequences in incentivized CEs is our main research question. Hence, we test whether having a different number of available alternatives changes choice behavior. The result of this hypothesis has important implications for the design of RCEs. If there is support for H2, practitioners of valuation studies would have to incentivize as many decisions as possible to ensure subjects exert the effort demanded by the experimental task.

**Hypothesis 3:** *The presence of incentives in CE changes search behavior.*

This study makes use of eye tracking technology. Previous literature has shown that time spent perusing a choice, and search behaviors within choice sets are predictors of decisions [Shi, Wedel,
and Pieters 2013). We evaluate whether the absence of economic incentives changes any of these measures with the eye tracking device. Support for H3 would imply that the cognitive processes of subjects are different in the presence of incentives than without them.

**Hypothesis 4a:** Higher numeracy increases the saliency of incentives in the CE.

**Hypothesis 4b:** Higher cognitive reflection increases the saliency of incentives in the CE.

Lipkus, Samsa, and Rimer (2001) show that numerical skills impact decision making. It has also been argued that cognitive reflection impacts the outcome of economic choices (Campitelli and Labollita 2010, Kahan 2013). One of our objectives is to gain insight into the mental processes that affect choice with and without incentives. In particular, we want to test to what extent the use of incentives to influence choice behavior is robust to variations in numerical skills and/or engagement with a task. In line with previous findings, higher numerical ability and/or being more attentive to a task should increase the performance in our induced value experiment. We employ biometric data – pupil dilation – to validate the scales of cognitive reflection and numeracy. Validating these procedures with non-intrusive biometrics, makes the results generalizable, as both numeracy and CRT can be implemented without any biometrics.

## 2 Related literature

### 2.1 Hypothetical bias and mitigation techniques

Within stated preference methods, CEs have become one of the most widely used tools in market and non-market valuation (Hess, Hensher, and Daly 2012). In spite of that, CEs face strong criticism due to several implied assumptions regarding the behavior of decision makers (Hensher, Rose, and Greene 2005) that could bias the results obtained from CEs. These biases can impact greatly the consequences and recommendations emanating from valuation studies (Colombo, Glenk, and Rocamora-Montiel 2016).
One of the key assumptions is that when responding to CE, subjects have preferences consistent with their true preferences (Sælensminde 2002, Ding, Grewal, and Liechty 2005). In other words, CEs assume that subjects reveal their true preferences when making their choices. There is ample evidence that this not the case, as eliciting preferences under hypothetical conditions results in hypothetical bias. List and Gallet (2001) report an average overstatement factor of 3.16 in willingness-to-pay elicitations in hypothetical versus incentivized conditions. Murphy et al. (2005) performed a meta-analysis on hypothetical bias using 28 studies and showed that there was a factor of 1.35 overstatement error in hypothetical methods compared to non-hypothetical methods.

Several procedures have become general practice to mitigate hypothetical bias, e.g., using calibration techniques prior to the choice task (Murphy et al. 2005). One of the most popular calibration techniques is cheap talk. Cummings and Taylor (1999) introduced the idea of communicating in a non-binding way with subjects before a hypothetical choice task scenario. The findings in the literature on the effectiveness of cheap talk in reducing hypothetical bias are generally mixed, with some showing it reduces hypothetical bias while others find no effect (List 2001, Lusk 2003, Silva et al. 2011). Another calibration approach is to use honesty priming. In this case, prior to the hypothetical task, subjects are primed with statements that value honesty. The general idea is that participating in honesty priming tasks encourages individuals to become more truthful about their preferences (Bello and Abdulai 2016). This technique has been shown to reduce the magnitude of hypothetical bias in CE (de-Magistris, Gracia, and Nayga 2013).

While calibration techniques can be useful in mitigating hypothetical bias, the problem tends to generally persist. Vossler (2016), for example, noted that the use of cheap talk may even exacerbate the hypothetical bias problem since it can emphasize that choices are hypothetical. The most suitable solution for eliminating hypothetical bias in valuation studies is to use economic incentives (Brock and Durlauf 2001, Dong, Ding, and Huber 2010). It is to no surprise that valuation studies use incentivized decisions to elicit preferences from subjects (e.g. Lusk and Schroeder 2004, Michaud,
Llerena, and Joly 2013), with even better results in mitigation of hypothetical bias when subjects use their own money (Moser, Raffaelli, and Notaro 2014). In RCE, incentivizing is usually done by randomly choosing a binding choice set (Vossler, Doyon, and Rondeau 2012). This follows from the view in valuation mechanisms that having a non-zero probability suffices to have incentive-compatibility (Collins and Vossler 2009), and that it would not only be costly to make every choice consequential, but it could give rise to complementarities, wealth effects, and strategic behavior (Knetsch 1989, Yang, Toubia, and de Jong 2018).

A challenge in the use of incentives is that in many instances, not all the product alternatives presented to participants are physically available. In some cases, certain products or features of existing products are not available in real markets, or they do not even exist (Hoyos 2010, de Bekker-Grob, Ryan, and Gerard 2012). The question is what to do when some of the alternatives are not physically available, making it impossible to enforce the market institution. An important methodological implication then is whether respondents’ choice behavior changes based on the number of product alternatives eligible for purchase in an RCE setting. The answer to this question presents a test to the experimental design, since non-disclosure of whether all product alternatives are available when evaluating actual market goods could potentially be deceptive.

To explore how a treatment impacts choice behavior, using a design that allows for identification of optimal solutions a priori is necessary. One design tool that helps with this is the induced values (IV) framework, as shown by Taylor et al. (2001) who use it to study hypothetical bias in public good provision. The theory behind IV experiments is straightforward. In a controlled experiment the premise is that the decision between the options presented is determined by how these options differ. If the experimenter can unequivocally account for and measure how the alternatives presented are different, and which of the alternatives provides, for example, the highest monetary value, inference can be made on the behavior of participants assuming they make utility maximizing decisions (Smith 1976). Subjects in IV experiments are given choices that are identical
except for the rewards and costs assigned exogenously to these options by the experimenter, e.g. tokens with different values, shapes with different prices, etc. Therefore, the most convenient feature of IV experiments is that participants do not have private values for the presented goods. The values are determined by the properties of the goods assigned by the experimenters. In their study, Taylor et al. (2001) find hypothetical bias in voting patterns between a fully binding and a completely hypothetical condition for induced value goods. The present study explores the effects on choice within the IV framework of probabilities lying between the completely hypothetical and fully binding conditions.

2.2 Aspects impacting the effectiveness in the use of incentives

In most economic experiments, treatment effects manifest if the treatments are salient for subjects (Bardsley 2005). The literature has shown that when subjects fail to correctly identify which game is being played, their decisions are not connected to the consequences of their actions (Cason and Plott 2014). In a CE framework, the disconnection between actions and consequences can yield choices that are not reflections of true preferences. In particular, for valuation studies the value formation and value elicitation tasks must be salient for subjects so that the results of the experiment are demand revealing (Cerroni et al. 2018).

Our experimental design allows us to evaluate the connection (or disconnection) between actions and consequences by exogenously changing the saliency of incentives in the CE. This characteristic in our design distinguishes our study from previous work on altering potentially binding decisions. Closely related to the present study, is the work of Carson, Groves, and List (2014), who evaluate changes in the probability of enactment on a voting system for provision for a public good. In their study, the probability of a referendum being binding is altered exogenously by the researchers. The approach of the present study builds on the results of Carson, Groves, and List
(2014) by evaluating the effects not of the probabilities of enactment, but of the selected good being available. Another contribution of the present study is the use of individual incentives, which are theoretically different from group incentives which impacts private good valuation research — but not public good provision research directly.

Poe and Vossler (2011) highlight the relevance of consequentiality and incentives in economic research. Recall that shedding light on which mechanisms enable or inhibit the importance of incentives to subjects is another objective of our study. We obtain insights into how individual heterogeneity impacts the perceived saliency by evaluating the treatments across different characteristics of individuals. One dimension of individual heterogeneity that can potentially influence the results of CE and other value elicitation techniques is the cognitive ability of subjects (Alós-Ferrer et al. 2012). In particular, numeracy skills are one aspect of cognition that has drawn the attention of researchers. Numeracy can be broadly defined as “the ability to process basic probability and numerical concepts” (Peters et al. 2006, p. 407). The main argument why numeracy impacts valuation and decision making is that subjects cognitive ability with respect to numbers can impede or enable identification of the optimal course of action (Kløjgaard, Bech, and Søgaard 2012). Hence, subjects who possess low numeracy skills may be less likely to make the connection between their choices in a CE and the consequences of those choices.

The literature in economics, finance, and other related fields shows that under laboratory and field settings, numerical skill plays an important role in decision-making (Robinson 1998, Banks, O’Dea, and Oldfield 2010, Chen et al. 2012). Bias in responses due to low numeracy skills can be a big problem in experimental and non-experimental settings since a large proportion of people even in developed countries have low numeracy skills (Lusardi 2012). Numeracy skills can be assessed using different techniques (Weller et al. 2013). Subjective measures, like the ones developed by Fagerlin et al. (2007), carry the benefit of measuring numerical ability without imposing a higher cognitive load on subjects. Their main shortcoming, however, is that they are self-reported measures and do not
necessarily correlate with real numerical ability (Dunning, Heath, and Suls 2004). In our study, we employ an objective measure for capturing numerical ability (Burkell 2004). Specifically, we adapt to our context the numerical ability questions developed by Schwartz et al. (1997).

Subjects in CE could also fail to recognize the saliency of incentives because they are simply not paying attention, unengaged. Therefore, another potential source of deviations from optimal behavior in valuation experiments is the level of attention to the choice task (Hensher 2006). There is evidence, for example, that subjects may systematically attend or ignore information in CE (Hensher, Rose, and Greene 2012, Scarpa et al. 2013, Hole, Kolstad, and Gyrd-Hansen 2013). If subjects are not paying attention, the incentive structure is likely not salient to them, implying that the consequences of their actions are not clear, and choices do not necessarily reflect their true preferences.

According to Grebitus, Lusk, and Nayga (2013), participants who do not pay attention during the entire choice task share some personality traits, and accounting for these subjects can explain a large portion of hypothetical bias. A popular method that measures engagement is the cognitive reflection test (CRT). The test was designed by Frederick (2005) to gauge subjects ability to suppress an intuitive and spontaneous, but ultimately wrong, answer in favor of a reflective and deliberative right answer. The test is simple and easy to implement. It consists of three questions. A higher score indicates a higher reflective state; in other words, a higher degree of engagement to the task at hand, which does not necessarily imply a higher level of cognitive ability (Hoppe and Kusterer 2011). We use CRT as an indicator of the level of engagement of participants to the choice tasks and evaluate the treatment effects for each level.
2.3 Eye tracking measurement

The use of biometrics enables researchers to explore other elements in the behavior of decision makers. In this article, we combine traditional methods with advances in biometrics to provide a more comprehensive picture of the dynamics in the behavior of economic choices with different saliency of incentives. Biometric technology like eye tracking allows a non-invasive measurement of subjects' performance. In particular, eye tracking evaluates gaze fixations of participants while responding to stimuli on a computer screen. The fovea is the portion of the retina responsible for visual information processing, projecting only about 2% of the visual field. The eye tracker captures the movement between stimuli to allow focus of the fovea in order to process new information (Duchowski 2003). Eye tracking devices are basically a set of high resolution infrared cameras. These cameras follow the subjects' eyes and gather their position on the computer screen, distance to the screen and, depending on the device, other measures related to visualization.

Using eye tracking in economics is not new, and it is gaining traction as the technology becomes more accessible. Eye tracking contributions to the literature span across different decision-making aspects. For example, Louviere (2006) shows with eye tracking that order and fatigue effects matter in CE, while Rasch, Louviere, and Teichert (2015) use it to measure different affective states leading to purchasing. Eye tracking has also been useful to monitor and model systematic non-attendance (Balcombe, Fraser, and McSorley 2015, Chavez, Palma, and Collart 2017).

Maughan, Gutnikov, and Stevens (2007) find that more time spent on an alternative increases the likelihood of selection. Eye tracking has also been utilized to model search behavior (Krajbich, Armel, and Rangel 2010, Grebitus, Roosen, and Seitz 2015, Khachatryan et al. 2017, Van Loo et al. 2018), use search patterns as a predictor of choices (Van der Lans, Pieters, and Wedel 2008), and explore top down or bottom up decision making (Orquin and Mueller Loose 2013). This study uses eye tracking to explore how visual search changes with economic incentives. In particular, we
measure the differences on time spent by choice set in the presence of incentives. We contribute to the extant literature by comparing search behavior across our treatments.

Eye trackers also provide measures of pupil dilation, a good indicator of attention and engagement (Wang, Spezio, and Camerer 2010). We use eye tracking technology, in addition to CRT, to evaluate attention and engagement between treatments. We use pupil dilation as a direct measure of attention and engagement, and test for heterogeneity in the treatment effects across subjects with different levels of eye tracking measures, numeracy skills and CRT scores.

3 Methodology

3.1 Experimental design

The experiment was conducted with general population subjects (i.e., non-students) recruited through local newspaper ads in a small city in southern United States. A total of 152 subjects participated in the experiment. Subjects were randomly assigned to one of four treatments.

We designed an induced value (IV) choice experiment where the optimal strategy of agents is profit maximization. To create our design, we use the shape/value concept developed in the IV experiment of Luchini and Watson (2014), but with different shapes and values. Specifically, each subject was presented with twelve choice sets, each consisting of two alternatives and an opt-out status quo option for none of the available alternatives. Each alternative consisted of a shape with predetermined values depending on three attributes: price ($0.5, $1.0, $1.5, $2.0), shape (square=$0.5, triangle=$1.0, circle=$1.5), and color (green=$0.5, blue=$1.0). A sample of the choice sets is presented in the appendix. The experimental design was developed in Ngene (ChoiceMetrics 2014) with the algorithm to maximize orthogonality and balance, and had a final D-error of 0.1492.
In the experiment, an agent maximizes profits by choosing the alternative with the highest benefit-cost differential. That is, participants can calculate the value of each shape based on its color and shape and subtract the purchasing price. They make profits if the price is lower than the value of the shape; however, only one of the alternatives maximizes profits. Every choice set had an alternative with nonnegative payouts. Half of the choice sets included one alternative with negative payout. Since attention and engagement are important considerations in our study, we include choices with negative payouts as a way to identify inattentive subjects. Selecting an option that yields a negative payout would indicate a subject was not paying attention, did not understand the game, or does not have the payout structure clear. This technique is in the same spirit as research that has used trap questions to identify unengaged participants (Malone and Lusk 2018). Our setup has the advantage that the negative payoffs are incorporated directly in the choice experiment and not included as instructions or separate questions.

With 12 choice sets, a total of 24 shape alternatives were presented to each participant. The experimental manipulation changes the number of alternatives eligible to become binding in each treatment. The number of potentially binding alternatives in each treatment is summarized in table 1. We avoid deception by using the following procedure. Each of the 24 alternatives was written down on a piece of paper. For each treatment, participants were informed that $n$ randomly selected alternatives would be placed inside a box to be eligible to become the binding product. The number of alternatives placed inside the box corresponds to the treatment assignment (0, 8, 16 or 24). The box containing the potentially binding products was placed next to the participant to ensure that the number of alternatives available to be binding was salient.

After reading the instructions, subjects completed a practice round. The results of the practice round were extensively discussed to ensure that participants understood the procedure. Once participants had no more clarification questions, and verifying they remembered and understood the payout structure, they advanced to the choice task stage. The experiment was incentivized in all
treatments except the purely hypothetical control. To incentivize subjects, we informed them that a bonus to their participation fee was at stake. The bonus was the profit they made during one randomly selected choice set. In order to determine the random choice set, each subject rolled a twelve-sided die. The number they rolled was the binding choice set and participants kept the profits they made in this round only if their chosen alternative was inside the box.

After completing the choice task, subjects filled a survey consisting of demographic questions, a numeracy skill quiz adapted from Weller et al. (2013), and the cognitive reflection test (Frederick 2005). Upon completing the survey, subjects rolled the 12-sided die to determine the binding choice set and were paid $20 for participating plus any profits made in the choice task, if applicable. These profits could be as much as $12 if all the choices are correct and all are enacted and as little as $0 if the dominated alternative or the opt-out option were chosen every time.

The experiment was presented as a slide show on a computer screen. The order of the choice sets was randomized to account for ordering effects. The highest paying alternative was the option on the left in half of the choice sets to balance the position presentation. A slide with a distractor at the center was presented for 3 seconds between each choice set to re-center the attention of subjects and avoid previous choice bias. Subjects could spend as much time as needed on the choice sets. All the data was collected in a laboratory with no windows, using fluorescent light of 750 lumens and 6500 Kelvin color temperature to light the room, thus keeping luminosity constant across subjects. We used the iMotions platform (iMotions 2016) to display the choice sets on a 1920 x 1200 pixels screen using a Tobii TX-300 screen based eye tracking device. The device was embedded to the computer screen while tracking and recording eye movements using near-infrared technology at a sampling rate of 120 data points per second. At the beginning of the experiment, the eye tracking device was calibrated to ensure proper data collection for each individual, using a nine-point calibration method.
To analyze the eye movements of subjects on the information presented we defined areas of interest (AOI) on the slides with the choice sets. These AOIs were polygons (rectangles) that enclosed the shapes, the price for each shape, and the opt-out alternative respectively. A total of five AOIs per choice set were constructed. Since the slides were built to be symmetric, the AOIs were also symmetric and equidistant. The gaze and fixation data from the AOIs was used for the analysis of eye tracking metrics of participants.

3.2 Theoretical Framework

Our analysis intends to quantify the magnitude of the effect of the different treatments on the number of optimal choices made by subjects in an induced value CE context, and to check for heterogeneity in the effects across subjects with different levels of numeracy skills and CRT scores. The latter provides valuable insights into the possible mechanisms behind the treatment effect results. In our analysis, we assume that agents have a monotonic utility function over the payouts and that they are utility maximizers. In other words, they prefer the option with the highest payout. With these assumptions, we evaluate choices in a random utility framework (McFadden 1974) to model the probability of subjects making the optimal profit-maximizing choice in each choice set. Since subjects make repeated selections across the choice sets, our data have a panel structure. To accommodate for the longitudinal dimension, we include $t$ in the model to render a random utility across time. In this framework the utility that individual $n$ receives from selecting option $j$ in choice set $t$ has the form of $U_{njt} = \beta x_{njt} + \varepsilon_{njt}$, where $\varepsilon_{njt}$ is an iid error term following an extreme value distribution, independent of $x_{njt}$ and uncorrelated to $n$ and $j$, while $\beta$ is describing the $n$-th individual with respect to which treatment the decisions were made under, subjects’ numeracy and CRT scores, and demographic information (gender, age, race, education level, and hourly income). With this setup
the probability of choosing the profit-maximizing alternative can be estimated with a panel logit, which is what we use for the estimation procedure.

4 Results

The sample consisted predominantly of white individuals (>65%) with an average age of 36 years old and mean yearly income of $64,500. Over half of the participants in each treatment were females and the majority had a college education. We conducted a balance test by comparing the demographics across treatments and found no statistically significant differences based on Mann-Whitney (MW) tests (p>0.10). The exception is gender where the 33% and 66% treatments had less females than the other two treatments. A summary of key variables is presented in table 2. Though the sample is balanced, it is not representative of the general population. The proportions of individuals across demographics does not match those of the US Census. The recruitment process may have had an influence in this regard, as using newspaper ads for recruiting might generate some selection bias in the sample. The objectives of the study, however, do not aspire to provide policy recommendations, but instead inform practitioners in market valuation. Under these conditions a representative sample is of lesser concern given the methodological nature of our research question.

The results are presented in three sections. First, we present the effects of the number of alternatives on the proportion of optimal choices in the IV experiment. Then, using biometrics we document differences in search dynamics, engagement and attentiveness with and without economic incentives. Finally, we assess the optimal choice behavior for individual level characteristics, numeracy and cognitive reflection scores across all treatments.

4.1 How much do incentives matter?

In order to position the findings of this study in the literature, as a starting point the existence of a difference between hypothetical and incentivized decisions is shown. Many of the CEs conducted are used to provide valuations for goods, where the subjects could use their choices as an opportunity to
signal policy makers (Nechyba and Strauss 1998). Since the setting for this study has no consequence on others or profit for the subject in the hypothetical treatment, differences between hypothetical and non-hypothetical treatments should be expected.

To measure the effect of economic incentives (Hypothesis 1), we first compare the proportion of optimal choices in each treatment (figure 1). The graph illustrates how the proportion of optimal choices is statistically lower when economic incentives are absent compared to any of the incentivized treatments (MW p<0.01). Incentivizing the CE increases the ratio of optimal profit-maximizing choices. The proportion of optimal profit maximizing choices for incentivized treatments were not statistically different between each other (though it may appear as the 33% treatment had a higher ratio). The average profits were $8.09 for the 33% treatment, $7.79 for the 66% treatment, $7.94 for the 100% treatment, and $6.75 in the 0% treatment. On average, the profits in the three incentivized treatments is 1.18 times higher than the hypothetical control.

[Insert figure 1 here]

Hypothesis 2 investigates whether the number of alternatives eligible to become binding changes the proportion of profit maximizing choices when incentives are used. We compare the mean percentage of optimal choices in each treatment ($60.2_{T_1}, 67.1_{T_2}, 65.1_{T_3},$ and $65.4_{T_4}$). The results indicate that the proportion of optimal choices in the incentivized treatments are not statistically different from each other (MW p>0.10), which does not support H2: the response to incentivized options is not monotonic. The results suggest that once incentives are used, the number of alternatives available for purchase does not change the proportion of optimal choices in our three incentivized treatments.

One explanation for this result is the engagement of subjects. Though most of the literature sampled implies that incentivizing is a binary choice – all or none – our results show that this need

1 We evaluate the individual derived utilities from the choices. No consistent pattern in the decision errors was identified, except for the opt-out option being selected more often in the hypothetical treatment.
not be the case. Subjects are treating the choices as if all the alternatives were potentially binding when only a third or two thirds of them are. In line with the findings of Yang, Toubia, and de Jong (2018) our study shows that the agents are not responding to the potential enactment of their choices in a monotonic manner, but instead in a zero versus non-zero bimodal manner.

To test the robustness of the unconditional analysis discussed above, we also conducted a conditional analysis using a random effects panel logit model on the probability of making optimal choices in each choice set (table 3). The variables used for estimation are treatment dummies and demographics. The parameter estimates and marginal probabilities of the model without demographic controls are shown in the first two columns of table 3. The last two columns of table 3 show the estimates and probabilities for the model controlling for demographics.

All three incentivized treatments increase the probability of making optimal choices with or without controlling for demographic heterogeneity. Specifically, in the model including demographic controls, relative to the 0% treatment, the probability of making an optimal choice is 7.43% higher for the 33% treatment, 6.33% higher for 66% treatment, and 7.21% higher for the 100% treatment. There is no statistically significant difference in the marginal probabilities of making optimal choices across the three incentivized treatments. This result is consistent with the unconditional analysis discussed above and provides further support to the notion that the improvement in performance granted by incentives is not increasingly monotonic; i.e., once incentives are used there is no need to have all the alternatives made available for purchase.

4.2 Behavior and engagement

To evaluate the behavioral processes and search dynamics across the treatments (Hypothesis 3), we use eye tracking measures. First, we compare total visit duration (TVD) across treatments. TVD is the amount of time subjects spent looking at the alternatives in each choice set. This eye tracking measure is strongly advised against when comparing two or more different stimuli, as the differences in TVD
between stimuli need not be because of the treatment variables, but a result of inherent differences between stimuli (Orquin and Holmqvist 2018). In our study, however, the stimuli (choice sets) are identical across treatments. Therefore, we use TVD to measure whether subjects are willing to invest more time evaluating the choice sets with different levels of incentives. We found no statistical differences in the average time spent on the alternatives across treatments ($1.79_{T1}, 1.85_{T2}, 1.74_{T3},$ and $1.86_{T4}$ $MW$ $p>0.10$). This indicates that incentivizing the decisions does not make subjects spend more time evaluating the choice set. Furthermore, the lack of statistical differences in time spent perusing the choice sets indicates that if subjects are more engaged with the decision-making process – as it will be discussed in the next section – it is through a cognitive process, not an information gathering process.

To explore any other behavioral differences between treatments, a TVD variance ratio test (Brown and Forsythe 1974) was conducted. This test revealed that the variance of TVD is larger for the non-incentivized group compared to any of the incentivized treatments ($p<0.05$). In contrast, the variance of TVD is not statistically different between any of the incentivized treatments ($p>0.10$). These results suggest that, on average, subjects behave similarly in terms of the time they spend evaluating each alternative whether incentives are present or not. However, in the hypothetical case, search patterns are more erratic, having a wider distribution (i.e., subjects spend far more time than the average or far less). One possible explanation for this finding is that some participants have an inherent desire to engage while others do not (Hamilton, Nickerson, and Owan 2003). In the absence of economic incentives, naturally engaged subjects will spend more time perusing the choice sets, while the unengaged subjects will breeze through it. In the presence of economic incentives, the commitment level across subjects increases, reducing the variation in the time spent on the choice task. This is in line with the findings described earlier. Incentivizing provides subjects with enough motivation to engage with the choice task.

---

2 To account for sample size per treatment, a 200-sample bootstrap was performed and produced the same results.
The second eye tracking measurement we evaluated is pupil size, which is used as an indicator of engagement with the task (Einhäuser et al. 2008). The eye tracking device uses the calibration data to estimate the pupil dilation in each recording, i.e. every 120 milliseconds in our study. Figure 2 shows average pupil dilation in millimeters for each treatment across all choice sets. We aggregate across choice sets as the variation within treatment across choice sets is not significant and the choice set order was randomized. Recall that all choice sets are identical. The only difference across treatments is the number of alternatives potentially binding. Thus, with equally complex stimuli across treatments, differences in pupil dilation, if there were, could be interpreted as the task deemed as more relevant by subjects (Orquin and Mueller Loose 2013). We find that pupil dilation is statistically higher for the incentivized treatments relative to the hypothetical treatment (MW p<0.01). This result serves as a biometric indicator of higher engagement in the incentivized conditions. It has been shown in the literature that in the presence of economic incentives, attention is enhanced (Small et al. 2005) even when subjects have no prior experience with the task (Heslin and Johnson 1992). In other words, with incentivized decisions, subjects pay more attention to the task. This is congruent with our findings. With pupil dilation as a measure of attention and engagement, we find evidence once more that incentives increase the engagement of subjects with the task, but that the level of incentives does not interact with how focused subjects can be.

4.3 Individual characteristics effect on the reaction to incentives

To discuss mechanisms through which the treatments generate different number of optimal choices, we first direct our focus to the numerical ability of individuals (Hypothesis 4a). The scale used to measure the numerical ability is an adaptation of the work of Weller et al. (2013). This scale measures the ability to understand and use numerical information, such as monetary amounts and probabilities. In their series of studies, the authors found the scale had robust predictive validity.
across diverse age and educational level groups. The complete scale consists of eight items, from which we selected five and replaced the rest with the cognitive reflection test (Frederick 2005). The distribution of the numerical scores was skewed to the right in every treatment. With this in mind we segregate the sample into individuals above the median and below the median, which did not differ across treatments. We classify the subjects who score below the median as low numerical ability subjects and those who score above the median as high numerical ability subjects.

We pool the results across all treatments to compare the number of optimal choices for subjects who had a score above the median and below the median (figure 3). Subjects who scored above the median in the numerical ability test made significantly more optimal choices than those who scored below the median (MW p<0.01). To evaluate the interaction of the numerical ability with the treatments, we plot the percentage of optimal choices for both groups of numerical ability for each of the treatments on figure 4. The results by treatment show that subjects with low numerical ability (left panel of figure 4) make less optimal choices regardless of the treatment, and the number of optimal choices is not statistically different between treatments for subjects with low numerical ability (MW p>0.10). This is the expected result, as the task at hand required manipulation of numbers to be able to identify optimal alternatives. It makes sense that subjects with low numerical abilities performed worse than their high numeracy counterparts across treatments. This result is also in line with the idea that numerical ability is necessary for incentives to be salient in economic experiments (Kløjgaard, Bech, and Søgaard 2012).

The next panel (right panel figure 4) shows that high numerical ability subjects react to incentives, i.e. the monotonic increase in performance that was proposed in H2 is seen for high numeracy subjects. The number of optimal choices for this group was statistically lower when no incentives were given than in any of the incentivized treatments. The highest number of optimal
choices achieved by these subjects occurs in the 100% treatment. The number of optimal choices in the 100% treatment is not statistically different from the number of optimal choices under the 66% treatment, but it is statistically higher than the number of optimal choices under the 33% treatment. The 33% and 66% treatment are not statistically different between each other. These results indicate that high numeracy individuals are able to exert more effort if the task has higher levels of incentives or decide to shirk if the incentives are not interesting enough for them. This is important to account for when conducting CE.

To further explore the interaction of the numerical ability with the treatments, we estimate a random effects logit regression on the likelihood of choosing the optimal alternative for low and high numerical ability groups separately. The estimation includes treatment dummy variables, the score in the CRT, with and without demographic controls. The estimation without demographic controls is shown in the first two columns of table 4, while the estimation with demographics is shown in the last two columns.

The first three rows in table 4 show marginal probabilities of the treatments. The results of the estimations are congruent with the unconditional analysis presented before. Subjects with low numerical ability do not react to incentives and the number of optimal choices is low regardless of the treatment. Subjects with high numerical ability on the other hand, do react to incentives and the probability of making optimal choices increases with the number of alternatives available. In the estimation with demographic controls, the probability increases by 10% in the 33% and 66% treatments and by 17% in the 100% treatment. This is an interesting result: there is an interaction of the natural ability of subjects to manipulate numbers with the probabilities of enactment of their choices. Thus, there exists a monotonic relationship of responsiveness to the probability of enactment, conditional on the numerical ability of subjects.

At this point we bring attention to another result. We find that reflective state; i.e. how much subjects care, measured through the CRT, can also explain the number of optimal choices. The
parameter estimate for CRT (fourth row in table 4) is positive and statistically significant in the estimations with demographic controls for both groups of numerical ability. This indicates that numerical ability is a good indicator of optimal choice making in the context of the experiment and so is attentiveness.

This last result leads the discussion into the second potential explanation for the observed choices: the reflective state of subjects (Hypothesis 4b). We captured subject attention by measuring pupil dilation (Hoeks and Levelt 1993, Wang, Spezio, and Camerer 2010) and reflective state with the CRT. To validate the CRT as an adequate measure of engagement, we evaluate the relationship between pupil dilation and CRT scores. We use the pooled data of all subjects to measure the average pupil dilation for each value of CRT score (figure 5). Based on the linear relationship that is observed in figure 5, we estimate a linear model of the pupil dilations as a function of the CRT scores across the different choice sets. We find that the estimated parameter is statistically significant, even when controlling for other variables. This indicates that CRT scores correlate with an objective, nonintrusive biometric measure of engagement. This result allows researchers to gauge engagement without the use of an eye tracking device by implementing the CRT, making it more widely applicable.

[Insert figure 5 here]

To use CRT scores for the cognitive reflection analysis, we examine the distribution of CRT scores in each treatment. We find no differences in the sample proportion of CRT scores between any of the treatments (Wald p>0.10). Given these features of the relationship between CRT and reflection, we segregate the entire sample based on level of attentiveness of subjects. Subjects with a score of three in the CRT were classified as highly attentive subjects. Those with a score of zero were considered low attentive subjects. The subjects with a score of one or two are classified as moderately attentive.

With this classification we explore the influence of engagement alone on the number of optimal choices by comparing the ratio of optimal choices by engagement levels (figure 6). Low
attentive subjects make significantly less optimal choices than moderate and high attentive subjects (MW p<0.01). The number of optimal choices of subjects with moderate and high attentive states are not statistically different (MW p>0.1).

[Insert figure 6 here]

To evaluate the interaction between the treatments and engagement we further explore the effects of attentiveness across the treatments in figure 7. The top left panel of figure 7 shows the proportion of optimal choices for low attentive subjects across treatments. The average proportion of optimal choices is around 57% for these subjects, which is statistically lower than the moderate and high attentive groups (MW p<0.01). Interestingly, there are no significant differences in optimal choice ratios across the four treatments for low attentive subjects (MW p>0.10). Hence, the presence of incentives, in terms of the product availability, did not impact the number of optimal choices made by subjects with low engagement. A direct implication of this result is that product availability in CEs does not matter for subjects who are not paying attention to the task, as it seems that no amount of product availability can incentivize them to provide truthful answers and engage with the task.

[Insert figure 7 here]

The bottom right panel of figure 7 shows the proportion of optimal choices for highly attentive subjects. The results for high attentive subjects show the opposite trend than the low attentive subjects. When subjects are attentive, the proportion of optimal choices is on average 74%, which is statistically higher than the average in the low and moderate groups (MW p<0.05). It is also notable that within high attention subjects, the optimal choice ratio was not statistically different across treatments (MW p>0.1). Put differently, for subjects with high engagement, the absence of incentives does not seem to affect the number of optimal choices they make. A consequence of this result is that using the data from subjects who are self-motivated and attentive, using the CRT as assessment tool, can improve the elicitation of truthful responses, even in the absence of incentives.
The other two panels in figure 7, the top right and bottom left, show the optimal choice ratios for moderately attentive subjects. Subjects with moderate levels of attentiveness did react to incentives and, more importantly, to the number of alternatives available to purchase. Within this group the number of optimal choices is an increasing function of the number of available alternatives, supporting H2. The ratio of optimal choices for moderately attentive subjects in the hypothetical treatment is 53%, which is statistically lower than any of the incentivized treatments (MW p < 0.01). Meanwhile, the proportion of optimal choices for the fully incentivized treatment is 81%, statistically higher than the two partially incentivized treatments (MW p < 0.01). The optimal choice ratios in the 33% and 66% treatments (70% and 72% respectively) are not statistically different (MW p > 0.10).

A couple of interesting observations across treatments and engagement levels can be made. The first one is that the optimal choice ratio of low attention subjects in any treatment and the optimal choice proportion of moderate engagement subjects in the hypothetical treatment are not statistically different (MW p > 0.1). Put differently, low engagement subjects with or without incentives yield the same outcomes in terms of optimal choices as moderate engagement subjects in a hypothetical treatment. This result indicates that for subjects with the lowest level of attentiveness, there is a disconnection between their actions and the consequences (i.e. if incentives are present, they are not salient regardless of the proportion of product availability, or they are not willing to exert any effort, at least for the treatments considered in this study). Subjects with a moderate level of engagement, however, responded to incentives by boosting their performance in the task; nonetheless, in the absence of economic incentives this group of subjects disengage and underperform.

The second observation is that the proportion of optimal choices of high attention subjects in the hypothetical treatment (first bar in the bottom right panel of figure 4) is not statistically different from moderately attentive subjects in any of the incentivized conditions (MW p > 0.1). This means that high attention subjects in a hypothetical task produce the same results as moderately attentive
subjects when the task is incentivized. The immediate implication of this finding is that for moderately attentive subjects, incentives matter to improve performance. In contrast, there seems to be no need to incentivize a CE for highly engaged subjects, since their performance in terms of optimal choices does not vary across the four treatments. Hence, the subjects that are focused on the task perform as well with or without incentives.

To further analyze the heterogeneity in results across level of engagement, we estimated random effects logit models for each group with and without demographic controls (table 5). In the estimation with demographic controls for the low engagement subjects, only the 100% treatment is marginally significant ($p < 0.10$), suggesting that the probability of choosing optimally is marginally higher in the 100% treatment than the other three treatments, ceteris paribus. For the highly engaged subjects, the probabilities of choosing optimally for the incentivized treatments are not significantly different from the 0% treatment. These results reinforce the idea that for highly attentive subjects, their performance is the same with or without incentives while for low attention subjects, a fully incentivized RCE may not improve their performance.

It is a different story, however, for the moderately engaged subjects, which seem to be driving the treatment effects found in the pooled model of Table 3. Specifically, the marginal effects for the incentivized treatments are all positive and statistically significant (i.e., all the incentivized treatments increased the probability of making the optimal choice). Furthermore, there are no statistical differences in the magnitude of the effects between the treatments. This result shows that subjects with moderate engagement levels react to incentives, but the probability of optimal choice behavior does not depend on the number of available alternatives used to incentivize the participants.

Another interesting result is that numerical ability is not statistically significant for any of the groups if demographics are accounted for. Contrasting this result with what was shown before; we find that although attention can explain variation in optimal choices across numeracy skills,
numerical skills do not have such explanatory power across engagement levels. In other words, accounting for engagement reduces the predictive power of numerical skills while attentiveness is a good predictor even when accounting for numerical ability.

In summary, the results from the conditional analysis suggest that the availability of all product alternatives in a RCE matters in terms of choosing the optimal choices for low engagement subjects, but that it does not matter as much for others. Interestingly, relative to the hypothetical treatment, the availability of 33% or 66% or 100% of the alternatives do not seem to matter at all for high attention subjects. The other interesting finding is that while incentives matter for the moderately attentive subjects relative to the completely hypothetical scenario, the percentage of available alternatives does not seem to matter for this group; i.e., there is not a significant difference in the magnitude of the effects between the three incentivized treatments. This result is in line with research showing that effort and attention compensate for lack of ability (Hau and Salili 1996). What we find in this study is an example of such behavior: subjects with high and medium engagement provide the best results, while subjects with low engagement perform the poorest, both irrespective of their cognitive numerical abilities.

5 Conclusions

To mitigate or reduce possible hypothetical bias in CE, the use of incentivized RCE is gaining popularity. RCE are normally implemented by making one of the decisions in a CE consequential (i.e., by randomly choosing a binding choice set). This situation becomes a challenge when not all the alternatives or attributes presented are physically available. In the absence of some of the product alternatives, practitioners can conduct a hypothetical CE or not communicate to subjects that some of the alternatives presented to them may not be available. The latter is done in an attempt to prevent subjects from perceiving the choice task as hypothetical, which can lead to disengagement with the task. Not informing participants that some of the alternatives may not be available for purchase,
would be considered deception by omission, so researchers should truthfully inform respondents about the number/proportion of product alternatives in the choice sets that are available in the experiment. But does the amount of product availability affect the saliency of the incentives in RCEs? That is, is there a minimum amount of product alternatives that are available in the experiment that would be considered salient enough by respondents to keep them incentivized to provide truthful responses and engaged with the task?

We conducted an induced value experiment where the optimal strategy of agents is to maximize profits, which in our design is achieved by selecting the optimal profit-maximizing alternative from each choice set. We varied the number of potentially binding alternatives between 0%, 33%, 66% and 100% in four different treatments. Our results suggest that having no consequences to the decision results in sub-optimal choices. The proportion of optimal choices in the purely hypothetical control was significantly lower than when economic incentives were present. However, if choices were incentivized, then having one-third, two-thirds or all the alternatives available in the RCE did not influence the results. The proportion of profit-maximizing choices was not statistically different across the three incentivized treatments (i.e., 33%, 66%, 100% treatments). This finding is important since it implies that one does not need to have all the product alternatives in a RCE be made available for the experiment to be salient enough to incentivize subjects to do the work, engage with the study, and provide truthful choices. Interestingly, based on our results, one could even conduct a RCE with just 33% of all product alternatives in the choice sets available for the study as a starting point.

Another important question addressed by this article relates to the behavioral aspects behind sub-optimal choices that could impact the perceived saliency of the economic incentives. Heterogeneity across participants reveals different effects of the number of alternatives on the optimal choice behavior. The results show that numeracy skills are not correlated to the number of optimal choices if subjects’ attention is accounted for. Measures of attentiveness using the pupil size
of respondents and CRT scores positively correlate with optimal choice behavior. In particular, subjects with low engagement levels do not react to economic incentives and have low performances across treatments. Participants with high levels of attentiveness, on the other hand, are not impaired by the lack of incentives and perform well even without economic incentives. It is the group of subjects with moderate attentiveness levels that react positively to all treatments, performing better with economic incentives than under hypothetical conditions. Additionally, since pupil dilation is positively correlated with CRT scores, CRT can be a useful tool to measure attentiveness without any biometric equipment, thus making it more generalizable. This also invites future research to review the robustness of these results in other datasets and contexts.

The results suggest that it does not matter in a RCE if the researcher has 33%, 66%, or 100% of product alternatives available in the study. One limitation of the study is that lower availability than 33% was not evaluated. Future studies should test the robustness of our findings with lower amounts of product availability (i.e., less than 33%). It is possible that there is a certain minimum threshold in terms of percentage of availability of products in RCE when the saliency of the incentives would disappear. Another important avenue for future research is extending the findings to market goods given that the current findings are based on a study using induced value goods, where a monotonic utility function for money is sufficient assumption to identify decision errors.
References


Research? Results from Surveys of Applied Experimental Economists and Students.”


iMotions. 2016. iMotions Biometric Research Platform 6.0. Copenhagen, Denmark: iMotions A/S.


Tables

Table 1: Summary of treatments

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Potentially binding alternatives</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% treatment</td>
<td>0</td>
<td>Hypothetical control</td>
</tr>
<tr>
<td>33% treatment</td>
<td>8</td>
<td>Partially incentivized - low</td>
</tr>
<tr>
<td>66% treatment</td>
<td>16</td>
<td>Partially incentivized - high</td>
</tr>
<tr>
<td>100% treatment</td>
<td>24</td>
<td>Fully incentivized</td>
</tr>
</tbody>
</table>
Table 2: Summary of means of demographics in the sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment 0%</th>
<th>Treatment 33%</th>
<th>Treatment 66%</th>
<th>Treatment 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average age</td>
<td>39.19</td>
<td>39.92</td>
<td>33.11</td>
<td>35.41</td>
</tr>
<tr>
<td></td>
<td>(2.32)</td>
<td>(1.60)</td>
<td>(2.51)</td>
<td>(2.77)</td>
</tr>
<tr>
<td>Mean yearly income (‘000s)</td>
<td>$65.91</td>
<td>$60.95</td>
<td>$67.19</td>
<td>$64.35</td>
</tr>
<tr>
<td></td>
<td>(7.01)</td>
<td>(7.27)</td>
<td>(7.97)</td>
<td>(7.01)</td>
</tr>
<tr>
<td>Females (%)</td>
<td>75.00</td>
<td>55.26</td>
<td>57.89</td>
<td>71.88</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>College degree (%)</td>
<td>100.00</td>
<td>97.44</td>
<td>92.11</td>
<td>96.88</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>White (%)</td>
<td>61.11</td>
<td>56.41</td>
<td>71.05</td>
<td>71.88</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Numeracy score (max 5)</td>
<td>4.00</td>
<td>4.15</td>
<td>3.84</td>
<td>3.31</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.16)</td>
<td>(0.19)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>CRT score (max 3)</td>
<td>0.94</td>
<td>1.36</td>
<td>1.05</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>Number of participants</td>
<td>36</td>
<td>39</td>
<td>38</td>
<td>32</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
**Table 3**: Random effects logit model of optimal choices

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Without demographics</th>
<th>With demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled</td>
<td>Marginal probabilities</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33%</td>
<td>0.372**</td>
<td>8.80%**</td>
</tr>
<tr>
<td></td>
<td>(0.151)</td>
<td>(3.54)</td>
</tr>
<tr>
<td>66%</td>
<td>0.289**</td>
<td>6.84%**</td>
</tr>
<tr>
<td></td>
<td>(0.136)</td>
<td>(3.20)</td>
</tr>
<tr>
<td>100%</td>
<td>0.258*</td>
<td>6.09%*</td>
</tr>
<tr>
<td></td>
<td>(0.133)</td>
<td>(3.15)</td>
</tr>
<tr>
<td>N</td>
<td>1596</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>2322.72</td>
<td></td>
</tr>
<tr>
<td>LogL</td>
<td>-1162.36</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Significance is indicated by *, ** and *** for the 10%, 5% and the 1% level or less respectively. Robust standard errors in parentheses. Marginal probabilities with respect to the 0% treatment.
Table 4: Random effects logit model of optimal choices by numeracy skill with and without demographic controls

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Without demographics</th>
<th>With demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33%</td>
<td>2.17%</td>
<td>9.06%**</td>
</tr>
<tr>
<td></td>
<td>(3.49)</td>
<td>(5.37)</td>
</tr>
<tr>
<td>66%</td>
<td>0.53%</td>
<td>12.63%***</td>
</tr>
<tr>
<td></td>
<td>(3.24)</td>
<td>(4.64)</td>
</tr>
<tr>
<td>100%</td>
<td>0.32%</td>
<td>18.92%***</td>
</tr>
<tr>
<td></td>
<td>(2.86)</td>
<td>(6.82)</td>
</tr>
<tr>
<td>CRT</td>
<td>0.283***</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>N</td>
<td>972</td>
<td>612</td>
</tr>
<tr>
<td>AIC</td>
<td>1487.54</td>
<td>814.43</td>
</tr>
<tr>
<td>LogL</td>
<td>-738.77</td>
<td>-402.21</td>
</tr>
</tbody>
</table>

Notes: Significance is indicated by *, ** and *** for the 10%, 5% and the 1% level or less respectively. Robust standard errors in parentheses. Marginal probabilities with respect to the 0% treatment.
Table 5: Random effects logit model of optimal choices by attention level with and without demographic controls

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Without demographics</th>
<th>With demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33%</td>
<td>3.00%</td>
<td>18.72%</td>
</tr>
<tr>
<td></td>
<td>(4.23)</td>
<td>(4.81)</td>
</tr>
<tr>
<td>66%</td>
<td>-0.19%</td>
<td>18.21%</td>
</tr>
<tr>
<td></td>
<td>(4.59)</td>
<td>(4.54)</td>
</tr>
<tr>
<td>100%</td>
<td>2.16%</td>
<td>23.45%</td>
</tr>
<tr>
<td></td>
<td>(3.22)</td>
<td>(9.05)</td>
</tr>
<tr>
<td>NUM</td>
<td>0.023</td>
<td>-0.057</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>N</td>
<td>792</td>
<td>552</td>
</tr>
<tr>
<td>AIC</td>
<td>1201.81</td>
<td>803.87</td>
</tr>
<tr>
<td>LogL</td>
<td>-595.90</td>
<td>-396.94</td>
</tr>
</tbody>
</table>

Notes: Significance is indicated by *, ** and *** for the 10%, 5% and the 1% level or less respectively. Robust standard errors in parentheses. Marginal probabilities with respect to the 0% treatment.
Figures

Figure 1: Optimal choices (%) by treatment
Figure 2: Pupil dilation (average of both eyes) by treatment
Figure 3: Optimal choices (%) by NUM scores above and below the median
Figure 4: Optimal choices (%) by NUM groups in each treatment
Figure 5: Average pupil dilation by CRT score.
Figure 6: Optimal choices (%) by CRT score
Figure 7: Optimal choices (%) by CRT score for each treatment
Appendices

Appendix A: For Online Publication

Experiment instructions:

The purpose of today's experiment is to help us understand purchasing decisions. To accomplish this purpose, you will be asked to select over a series of twelve items. I will explain how the experiment will work and we will have a practice round. After that you will choose from the alternatives presented and then we will ask you to fill in a survey. After the survey has been completed, you will receive payment for your participation in today's session.

The experiment we will conduct today will probably be different from any experiment you have had experience with previously. In each slide, we will present you with two shapes of different colors and you will have to choose which shape to buy. Each circle you will get $1.50, for each triangle you will get $1.00 and for each square you will get $0.50. As for the colors green will pay you $0.50 and blue will pay you $1.00. The value of each shape is the combination of its shape and its color. You will buy this shape and profit from the value of the shape. The price for each shape/color combination is shown underneath the shape. The profit comes from subtracting the price from the value the shape and color give.

(For the hypothetical treatment)
This is just an experiment, we just want to gather your preferences. If you decide you prefer one of the products, you will not pay the price and it will not affect your compensation.

(For the binding treatment with all products available)
This is a real experiment. All the alternatives shown are potentially binding. We will select one of the twelve rounds of colored shape pairs as binding and what you chose in that round will be your purchase. You will pay the stipulated price and in exchange you will receive the respective payout.

(For the binding treatment with not all products available)
This is a real experiment, but not all options of colored shapes shown are available for purchase. Right now, you will randomly draw [eight/sixteen] cards with the shapes in the experiment from a box and we will place them in the tumbler on the desk. These are unknown to you and to us until the end of the experiment when we will select one of the twelve rounds as binding. What you chose in that round will be your purchase. If the alternative you chose is one of the cards in the tumbler, i.e. part of the [33/66] percent available, you will pay the stipulated price and in exchange you will receive the payout. If it is not, you will not pay the
price and it will not affect your compensation.

Appendix B: For Online Publication
Sample choice set:

Option C
None of the Above