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

Daniel E. Chavez[1*], Marco A. Palma[1] and Rodolfo M. Nayga, Jr.[2]

[1] Department of Agricultural Economics, Texas A&M University, USA ; [2] Department of Agricultural Economics and Agribusiness, University of Arkansas;
[*]Corresponding author Email: danieleduardo11@tamu.edu

Invited presentation at the 2018 Southern Agricultural Economics Association Annual Meeting, February 2-6, 2018, Jacksonville, Florida

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

Daniel E. Chavez· Corresponding author. Research Assistant, Department of Agricultural Economics, Texas A&M University 2124 TAMU, College Station, Texas, 77843

(979)-587-4492 danieleduardo11@tamu.edu

Marco A. Palma Associate Professor, Department of Agricultural Economics, Texas A&M University. 2124 TAMU, College Station, Texas, 77843

(979)-845-5284 mapalma@tamu.edu

Rodolfo M. Nayga, Jr. Distinguished Professor and Tyson Endowed Chair, Department of Agricultural Economics and Agribusiness. University of Arkansas. Fayetteville, Arkansas, 72701

(479)-575-2299 Rnayga@uark.edu

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

ABSTRACT

The proneness of stated preference choice experiments to hypothetical bias has increased the popularity of incentivized or real discrete choice experiments (RDCE). One challenge that practitioners face when designing RDCE is that some of the product alternatives may not be available for the study. To avoid deception, researchers should truthfully inform respondents that only a certain percentage of the product alternatives is available for the experiment. But would the proportion of available products influence the results of the RDCE? Using an induced value choice experiment, we varied the number of potentially binding alternatives in four treatments: 0%, 33%, 66%, and 100% to assess the effect of availability of product alternatives on choice behavior. We designed the induced value experiment with a profit maximization optimal strategy for agents (i.e., with a unique known profit-maximizing alternative). Our results suggest that incentives matter in that the percentage of optimal choices was lowest in the 0% treatment. Interestingly, however, we did not find statistically significant differences in amount of optimal choices in the 33%, 66%, and 100% treatments, suggesting that one could conduct an incentivized RDCE without the need to have all the product alternatives be made available in the study.

Key words: Choice Experiments, Eye Tracking, Hypothetical bias, Induced values

JEL codes: C91, C18

"If men define situations as real, they are real in their consequences"
- *Thomas Theorem*
William & Dorothy Thomas, 1928

Discrete choice experiments (DCE) are widely used for market and environmental valuation. A well-established literature documents the recurring phenomenon of hypothetical bias in stated preference studies such as DCE (List and Gallet 2001, Murphy et al. 2005). That is, there is an overvaluation of products and services in the absence of economic incentives (Harrison and Rutström 2008). As a result, researchers are gravitating towards incentivized real discrete choice experiments (RDCE).

A challenge when designing RDCE is that some of the product alternatives presented in the choice sets may not be available yet in markets or physically present. After all, one important research question addressed by DCE 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, researchers using RDCE have to truthfully inform respondents that some of the products they are choosing from in the choice sets are not available.

Our research question is whether or not the number of products available to purchase in an RDCE affects the subjective perception of whether the experiment is hypothetical or real. In other words, we wish to know if there would be differences in choice behavior in RDCE when the number of product alternatives available for purchase differs. Put differently, is there a certain amount of product availability that is not salient enough anymore to incentivize subjects in RDCE? This is an important question given the challenge of coming up with all the product alternatives in a RDCE.

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, 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) a hypothetical case 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) a case where all alternatives are available for purchase. To avoid deception, we set up the experiment by randomly drawing a subset of the products 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 available for purchase when making their decisions. The full details of the procedure are described in the experiment design section.

The main objective of this article is to assess how the incentives for optimal profit maximizing behavior change with different number of products eligible to become binding in an RDCE. To do this, we first establish the magnitude of hypothetical bias (i.e., amount of deviation from optimal choices) in an induced value experiment. 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 level of cognitive reflection and attentiveness on the number of optimal decisions made by subjects as they may selectively ignore or attend information.

Consistent with the existing literature, we find that economic incentives matter. The proportion of optimal choices was significantly higher when economic incentives were used. Interestingly, however, if the experiment was incentivized with varying degrees of availability of products, we found no statistically significant differences in the proportion of optimal choices across the three incentivized treatments (i.e., 33%, 66%, 100% availability of products).

The results also show that while mathematical ability does not influence the number of optimal choices, cognitive reflection –which measures attentiveness– is positively correlated with optimal behavior. The number of optimal choices is higher for participants with the highest cognitive reflection test (CRT) scores. On the other hand, participants with the lowest CRT score behave the same regardless of whether economic incentives are used or not. To directly measure engagement across treatments, we also examined the choice behavior of subjects using an eye-tracking device to compare search dynamics and pupil size. The results suggest that CRT scores may be a useful tool to correct for potential hypothetical bias in non-incentivized DCEs.

2. Related literature

2.1. Hypothetical bias and mitigation

Stated preference methods have a prominent status in market and non-market valuation. DCE, in particular, have become one of the most widely used tools in stated preference (Hess, Hensher, and Daly 2012). However, DCE face strong criticism due to several assumptions regarding the behavior of decision makers (Hensher, Rose, and Greene 2005). One of the key assumptions is that when responding to DCE, subjects have preferences consistent with their true preferences (Sælensminde 1998). There is ample evidence that eliciting preferences under hypothetical conditions results in hypothetical bias. Murphy et al. (2005) performed a meta-analysis on hypothetical bias and using 28 stated preference studies showed that there was a factor of 1.35 overstatement error in hypothetical methods compared to non-hypothetical. List and Gallet (2001) report an average overstatement factor of 3.16 in willingness-to-pay elicitations in hypothetical versus incentivized conditions.

Several procedures have become general practice to mitigate hypothetical bias, e.g., using non-student populations, or more generally using a sample of the target population, and calibration techniques prior to the choice task (Murphy et al. 2005). For calibration in particular, several variations of the approach have been documented in the literature. 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 (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 DCE (de-Magistris, Gracia, and Nayga 2013).

While calibration techniques can be useful in mitigating hypothetical bias, the problem tends to generally persist. The most suitable solution for eliminating hypothetical bias is to use economic incentives by implementing RDCE (Brock and Durlauf 2001). This procedure is usually carried out by randomly choosing a binding choice set, i.e., to randomly select one (or more) choice sets as binding and enforce the market conditions based on the decision maker's choices (Vossler, Doyon, and Rondeau 2012). However, in many instances, not all the alternatives presented to participants are physically available. In some cases, certain products –or new 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 challenge to practitioners is what to do when some of the alternatives are not physically available, making it impossible to enforce the market institution. This situation brings up the question of whether this information needs to be communicated to respondents in order to avoid deception by omission. If this information is disclosed to participants, then an important methodological question is whether their behavior and choices would change based on the number of alternatives eligible for purchase in an RDCE setting.

2.2. Other aspects contributing to hypothetical bias

The cognitive ability of subjects is another dimension that can potentially influence the results of DCE and other value elicitation techniques (Alós-Ferrer et al. 2012). In particular, numeracy skills are one aspect of cognition that has drawn the attention of researchers. Numeracy is defined as the ability to understand and manipulate numbers (Peters et al. 2006). 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).

The literature in economics, finance, marketing 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 is that they are self-reported measures and do not necessarily correlate with real numerical ability (Dunning, Heath, and Suls 2004). 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).

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 that subjects may systematically attend or ignore information in DCE (Hensher, Rose, and Greene 2012, Scarpa et al. 2013, Hole, Kolstad, and Gyrd-Hansen 2013). According to Grebitus, Lusk, and Nayga (2013), there are some participants –who commonly share personality traits– that do not pay attention during the entire choice task. Accounting for these subjects who are not likely to state their true preferences can explain a large portion of hypothetical bias (Grebitus, Lusk, and Nayga 2013). One method to identify and isolate inattentive subjects is to use trap questions (Malone and Lusk 2018). Separating the results for attentive and inattentive subjects can be useful in analyzing and potentially mitigating hypothetical bias.

The cognitive reflection test (CRT), commonly used in the psychology literature, is gaining popularity in economic and marketing research. 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 attentiveness 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 attentiveness of participants to the choice tasks and a potential way to reduce the value gap due to hypothetical bias.

2.3. Eye tracking measurement

Eye tracking technology can be useful to evaluate visual attention, engagement, and search dynamics of economic choices. The popularization of eye tracking enables a non-invasive exploration of the behavior of decision makers. 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 the 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 subject's eyes and gather their position on the computer screen, distance to the screen and, depending on the device, other measures such as pupil dilation and luminosity levels.

Using eye tracking in economics is not new, but it is gaining traction as the technology becomes more accessible. Eye tracking contributions to the literature span across different decision-making aspects. Maughan, Gutnikov, and Stevens (2007) find that more time spent on an alternative increases the likelihood of selection. Louviere (2006) uses eye tracking to show that order and fatigue effects matter in DCE.

Eye tracking has also been used to model search behavior, predicting choices (Krajbich, Armel, and Rangel 2010, Grebitus, Roosen, and Seitz 2015, Khachatryan et al. 2017, Loo et al. forthcoming) and to measure different affective states leading to purchasing (Rasch, Louviere, and Teichert 2015). It has also been useful to monitor systematic non-attendance (Balcombe, Fraser, and McSorley 2015).

In addition to eye movement, eye trackers also provide other measures such as pupil dilation. Pupil dilation is a good indicator of attention and engagement (Wang, Spezio, and Camerer 2010). We propose to use pupil dilation as a direct measure of attention and engagement, and relating it to the evaluation of treatment effects, numeracy and CRT outcomes. 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.

3. Methodology

3.1. Experimental design

The experiment was conducted with general population subjects (i.e., non-students) recruited through local newspaper ads. A total of 152 subjects participated in the experiment. Subjects were randomly assigned to one of four treatments, with around 38 participants in each treatment cohort. The data from eight participants were incomplete and removed from the sample.

We designed an induced value (IV) choice experiment where the optimal strategy of agents is profit maximization. We use a simplified version of the IV experiment in Luchini and Watson (2014). Each subject was presented with twelve choice sets consisting of two alternatives and an opt-out *status quo* option for none of the available alternatives. Each alternative consisted of a polygon with predetermined values depending on three attributes: *price* ($0.5, $1, $1.50, $2), *shape* (square=$0.5, triangle=$1.0, circle=$1.5), and *color* (green=$0.5, blue=$1.0). The experimental design was developed in Ngene (ChoiceMetrics 2014) with the algorithm to maximize orthogonality and had a final D-error of 0.1492. A 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 our experiment, an agent maximizes profits by choosing the alternative with the highest benefit-cost differential. That is, participants can calculate the value of each polygon based on its color and shape and subtract the purchasing price. They make profits if the price is lower than the value of the polygon; however, only one of the alternatives maximizes profits. Nearly all the alternatives were nonnegative; only two choice sets included one alternative with negative profits. Note that selecting a negative alternative would yield similar results as using trap questions to identify participants not paying attention (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 IV polygon alternatives were presented to each participant. The experimental manipulation changes the number of alternatives eligible to become binding in each treatment. Participants were randomly assigned to one of the following conditions:

(1) 0 (0%) – Hypothetical Control

(2) 8 (33%) – Partially Incentivized-Low

(3) 16 (67%) – Partially Incentivized-High

(4) 24 (100%) – Fully Incentivized

We avoid deception by using the following procedure. Each of the 24 alternatives was written down in a piece of paper. For each treatment assignment, participants were informed that *n* randomly selected alternatives would be placed inside an urn 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 urn containing the potential 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, 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 urn.

After completing the choice task, subjects filled a survey consisting of demographic questions, a numeracy skill quiz based on the work of Schwartz et al. (1997) 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 and any profits made in the choice task, if applicable.

The experiment was presented on a 1920 x 1200 pixels screen of a Tobii TX-300 eye tracking device using the iMotions platform (iMotions 2016). 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.

3.2. Research hypotheses

In order to formally present our research questions, we propose the following hypotheses:

3.2.1. Hypothesis 1: Economic incentives matter in DCE.

The literature has extensively documented hypothetical bias in DCE. Our design allows measuring the effect of number of available product alternatives on choice behavior and hypothetical bias in a controlled environment by using an induced value setting.

3.2.2. Hypothesis 2: There are no differences across incentivized treatments.

As discussed earlier, one of the motivations of this article is understanding the behavioral and value consequences of incentivizing DCEs using various degrees of availability of the product alternatives in the choice task. Hence, we test if having a different number of available product alternatives changes the dynamics of choice. The result of this hypothesis has important implications for the design of RDCEs.

3.2.3. Hypothesis 3a: Numeracy predicts optimal profit maximizing behavior in the DCE.

3.2.4. Hypothesis 3b: Cognitive reflection predicts optimal profit maximizing behavior in the DCE.

It has been shown in the literature that numerical skills impact decision making (Lipkus, Samsa, and Rimer 2001). 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 numerical skills or other simpler ways to assess attentiveness to a task influence choice behavior. We employ biometric data, including eye tracking and pupil size to validate behavioral responses related to attentiveness and engagement. However, after validating these procedures, both numeracy and CRT can be implemented without any biometrics, thus making them more generalizable. If hypotheses 3a and 3b are rejected, and a numeracy and/or cognitive reflection effect is found to be a good predictor of rational economic behavior, our assumption is that it would be a positive effect. We believe that in line with previous findings, higher numerical ability and/or being more attentive to a task should increase the likelihood of choosing optimally in our induced value DCE.

3.3. Theoretical framework

Our analysis intends to quantify the magnitude of the effect of the different treatments, numeracy skills and CRT scores in the probability of making optimal economic choices in an induced value DCE context. To do this 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 model choices in a random utility framework (McFadden 1974). Since subjects make repeated selections across the choice sets, this provides a panel structure for the data. To accommodate for this 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 ε_{njt} is an iid error term following an extreme value distribution, independent of x_{ijt} and uncorrelated to *n* and

j, while β is describing the n-th individual with respect to the treatment, numeracy and CRT. The probability of choosing the profit-maximizing alternative can be modeled with a panel logit:

(1) $$P_{j|J} = \frac{\exp(U_{ijt})}{\sum_{j \in J} \exp(U_{ijt})}$$

4. Results

The sample consisted predominantly of white individuals (>65%) with an average age of 36 years old and a mean yearly income of $64,500. Over half of the participants in each treatment were females and the large majority had a college education. We conducted a balance test by comparing the demographics across treatments and found no statistical differences in Mann-Whitney (MW) tests (p>0.10). These results imply that differences across treatments are not driven by selection bias. A summary of key variables is presented in Table 1.

The results are structured in three sections. First, we present the effects of hypothetical bias in the IV experiment. Then, using biometrics we document differences in search dynamics, engagement and attentiveness with and without economic incentives. Next, we compare the outcomes across treatments. Finally, we assess the potential sources of hypothetical bias by looking at individual level characteristics, numeracy and cognitive reflection scores.

Table 1 Summary of demographics in the sample

	Treatment			
Variable	0%	33%	66%	100%
Average age	39.19 (2.32)	39.92 (1.60)	33.11 (2.51)	35.41 (2.77)
Mean yearly income ('000s)	$65.91 (7.01)	$60.95 (7.27)	$67.19 (7.97)	$64.35 (7.01)
Females (%)	75.00 (0.07)	55.26 (0.08)	57.89 (0.08)	71.88 (0.08)
College degree (%)	100.00 (0.00)	97.44 (0.03)	92.11 (0.04)	96.88 (0.03)
White (%)	61.11 (0.08)	56.41 (0.08)	71.05 (0.07)	71.88 (0.08)
Numeracy score (max 5)	4.00 (0.18)	4.15 (0.16)	3.84 (0.19)	3.31 (0.25)
CRT score (max 3)	0.94 (0.20)	1.36 (0.19)	1.05 (0.19)	0.825 (0.18)
Number of participants	36	39	38	32

Standard errors in parentheses

4.1. How much do incentives matter?

To measure the effect of economic incentives (Hypothesis 1) we first compare the proportion of optimal choices in each treatment (Figure 1). The graph show that 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 DCE increases the ratio of optimal profit-maximizing choices. Expected average profits were $8.09 for the incentivized-low, $7.79 for the incentivized-high, $7.94 for the fully incentivized and $6.75 in the hypothetical setting. On average the payout in the incentivized treatments is 1.18 times higher than the hypothetical control.

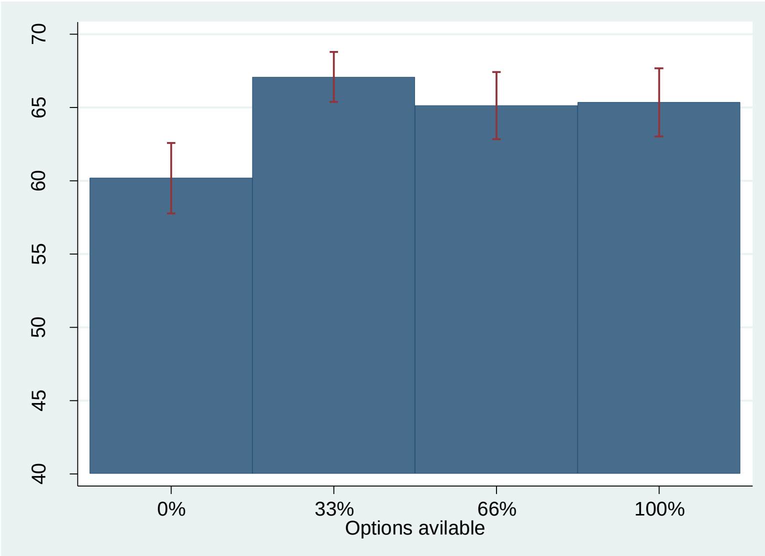

Figure 1: Optimal choices (%) by treatment

Hypothesis 2 investigates when incentives are used, does it matter how many alternatives are eligible to become binding? The proportion of optimal choices in the incentivized treatments are not statistically different from each other (MW p>0.1). This result shows 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.[1,2]

To evaluate the behavioral process and search dynamics in the DCE, 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. There are no statistical differences in the average time spent on the alternatives across treatments (MW p>0.10). However, a variance ratio test (Brown and Forsythe 1974) revealed that the variance is larger for the non-incentivized group compared to the any of the incentivized treatments (p<0.05). In

[1] In our study the lowest number of available alternatives was 33%. Future work may focus on lowering the number of alternatives to further test how that would impact choice behavior.

[2] We also run an ANOVA on a 200 bootstrap sample simulation of the results. In such test we found the within group variation of the incentivized treatments is not statistically different (p>0.10) but that the between group variation with the hypothetical treatment is statistically different (p<0.10).

contrast, the variance of WTVD is not statistically different between any of the incentivized treatments (p>0.10). This indicates that under incentivized conditions there is less variation in the time subjects spend evaluating each alternative.

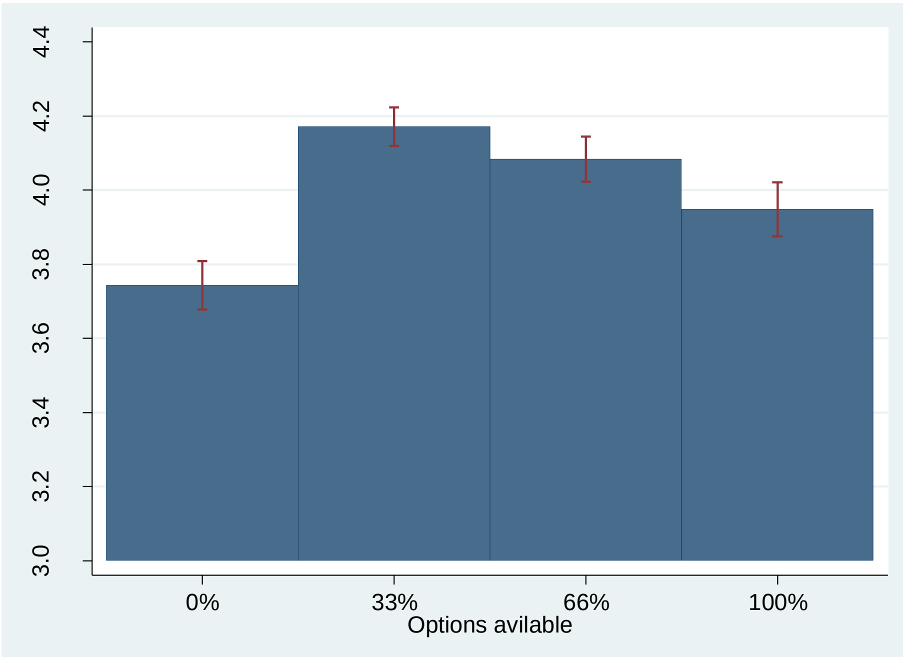

Figure 2: Pupil dilation (average of both eyes) by treatment

The second metric obtained from the eye tracker is pupil size, which is used as an indicator of engagement with the task (Einhäuser et al. 2008). Figure 2 presents the average pupil dilation in millimeters for each treatment. 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.

4.2. Identifying the effect of individual level characteristics on the reaction to incentives

Individual level characteristics may be useful to explain reactions to economic incentives. Recall that the sample was balanced across treatments. We also test for the distribution of CRT scores in each treatment. We find no differences in the proportion of CRT scores between any of the treatments (Wald p>0.10). Due to this feature, we compare the ratio of optimal choices by CRT score (Figure 3).

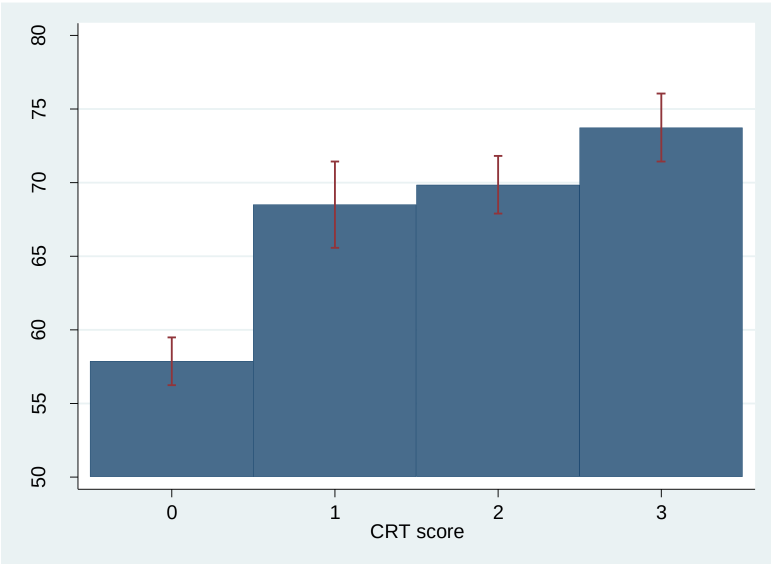

Figure 3: Optimal choices (%) by CRT score

It is shown in Figure 3 that low attention subjects (a score of zero in the CRT) make significantly less optimal choices (MW p<0.01). The number of optimal choices of subjects with medium (CRT scores between 1-2) and high attention (CRT score of 3) are not statistically different (MW p>0.1). Additionally, we explore the effects of the attention level captured by the CRT with and without economic incentives in Figure 4.

The top left panel of Figure 5 shows the proportion of optimal choices for the low attention subjects across treatments. The average proportion of optimal choices is around 57% for this group, which is statistically lower than the higher attention groups (MW p<0.01). It is important to note that for low attention subjects the hypothetical treatment and all the incentivized treatments yield optimal choice ratios that are not statistically different (MW p>0.10). In other words, the presence of incentives did not impact the number of optimal choices made by subjects with low attention.

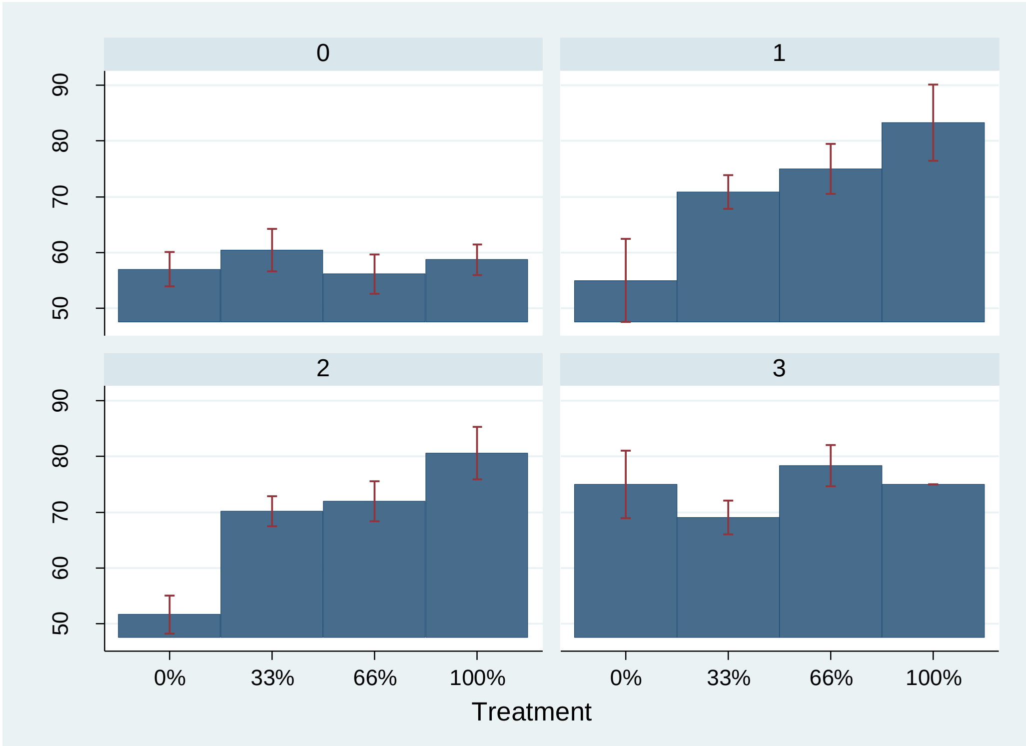

Figure 4: Optimal choices (%) by CRT score for each treatment

The bottom right panel of Figure 4 shows the proportion of optimal choices for the high attention subjects. The results for high attention subjects show the opposite trend. When subjects are focused, the proportion of optimal choices is on average 74%, which is statistically higher than average of any of the other attention groups (MW p<0.05). It is also notable that for high attention subjects only the incentivized-low treatment had a optimal choice ratio that was statistically lower than the rest of the treatments (MW p>0.1). Put differently, for focused subjects the number of optimal choices they make does not seem to be affected by economic incentives.

The other two panels in Figure 4, the top right and bottom left, show the optimal choice ratios for medium attention subjects. Subjects with medium 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. The ratio of optimal choices for medium attention subjects in the hypothetical treatment is 53%, which is statistically lower than the 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 partially incentivized treatments (MW $p<0.01$). The optimal choice ratios in the partially incentivized treatments, 70% and 72% respectively, are not statistically different (MW $p>0.10$).

A couple of interesting observations across treatments and attention 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 medium attention subjects in the hypothetical treatment are not statistically different (MW $p>0.1$). Put differently, low attention subjects with or without incentives yield the same outcomes as medium attention subjects in a hypothetical treatment.

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 medium attention 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 medium attention subjects when the task is incentivized. The immediate implication of this finding is that for medium attention subjects incentives matter and improve performance. For high attention subjects, in contrast, there is no need to incentivize a DCE. The subjects that are focused on the task perform as well with or without incentives.

This result implies that CRT scores can be a valuable tool to separate the attention levels of participants. In our experiment, using high attentive subjects – those with the highest CRT scores – produce identical results across treatments. CRT may prove a useful mechanism to mitigate hypothetical bias in situations where providing economic incentives is not feasible.

In order to support the use of CRT as an adequate measure of engagement and attentiveness, we compare the average pupil dilation of subjects by CRT scores (Figure 5). The engagement of subjects, measured by dilation of their pupils is higher for subjects with higher CRT scores (MW p<0.01). This indicates that CRT scores correlate with an objective, non-intrusive biometric measure of engagement captured by pupil dilation. This result is further explored in the following section using a panel logit specification.

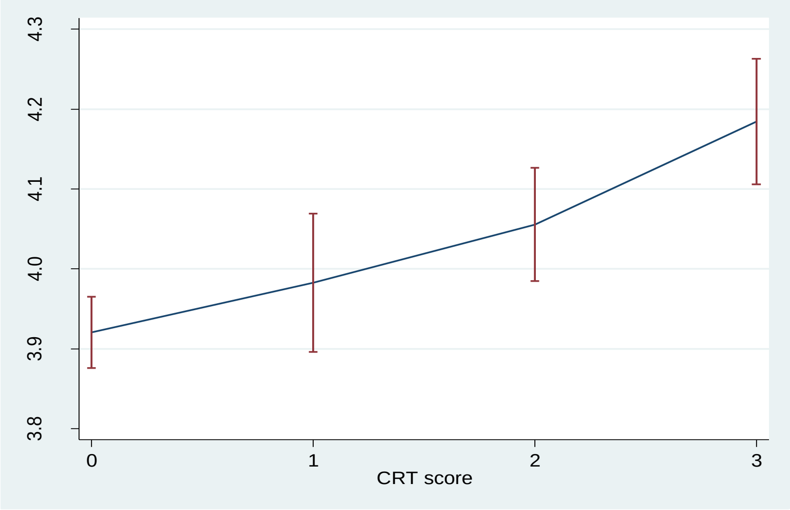

Figure 5: Average pupil dilation by CRT score.

A random effects panel logit was estimated to show the treatment effects on the probability of making optimal choices (Table 2). The variables used for the estimation include treatment dummies, numeracy scores, CRT scores, and demographics. All three treatments increase the probability of making optimal choices. Furthermore the coefficients for the incentivized treatments are not statistically different. This result provides further support to the notion that once incentives are used, there is no need to have all the product alternatives made available in the study.

Table 2. Random effects logit model of optimal choices

Parameter	Coefficient
TREATMENT	
33%	0.296**
	(0.147)
66%	0.292**
	(0.149)
100%	0.419***
	(0.154)
NUM	0.034
	(0.050)
CRT	0.261***
	(0.060)
AGE	-0.006
	(0.005)
EDUC	0.084
	(0.090)
FEMALE	0.296**
	(0.116)
RACE	-0.062
	(0.043)
HOURLY INCOME	-0.005*
	(0.003)
CHILDREN	-0.058
	(0.091)
MARRIED	-0.314**
	(0.125)
HOUSEHOLD SIZE	-0.045
	(0.057)
EMPLOYMENT	0.014
	(0.044)
AIC	2093.89
Log-Likelihood	-1035.94

Note: Significance is indicated by *, ** and *** for the 10%, 5% and the 1% level or less respectively. Robust standard errors in parentheses.

The coefficient for CRT score is positive and statistically significant. The sign and magnitude of the parameter imply that the scores in the CRT are a good predictor of the likelihood of choosing the optimal alternative. In contrast, the parameter estimate for numeracy is not statistically different from zero. From these results, we conclude that numerical ability is not necessarily relevant for optimal rational decision making in this context if we control for attention levels.

5. Conclusions

Due to the prevalent hypothetical bias in DCE, incentivized RDCE are gaining popularity. RDCE are implemented by making one (or several) of the decisions (i.e., choice sets) in a DCE consequential. 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 could conduct a hypothetical DCE 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. Not informing participants that some of the alternatives may not be available for purchase, however, could be considered deception by omission. To avoid deceiving respondents, the proper way of approaching this issue is for researchers to truthfully inform respondents about the number of product alternatives in the choice sets that are available in the experiment. But what is the minimum amount of product alternatives that are available that would be considered salient enough by respondents to keep them incentivized to provide truthful responses?

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.

In line with an extensive literature, our results show that having no consequences to the decision results in hypothetical bias. 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 to become binding 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 RDCE be made available for the experiment to be salient enough to incentivize subjects to provide truthful choices. Interestingly, based on our results, one could even conduct a RDCE with just 33% of all product alternatives in the choice sets available for the study.

Another important question addressed by this article relates to the behavioral aspects behind hypothetical bias. The results show that low numeracy skills are not correlated to the size of hypothetical bias. However, subject's attentiveness, measured directly using the pupil size of respondents and by the CRT scores positively correlates with optimal choice behavior. The pupil dilation is higher when incentives are present. Additionally, pupil dilation is positively correlated with the CRT scores. This result suggests that CRT can be a useful tool to measure attentiveness without any biometric equipment, thus making it more generalizable.

While our results suggest that it does not matter in a RDCE if the researcher has 33%, 66%, or 100% of product alternatives available in the study, 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 threshold in terms of percentage of availability of products in RDCE when the saliency of the incentives would disappear.

References

Alós-Ferrer, Carlos, Đura-Georg Granić, Fei Shi, and Alexander K. Wagner. 2012. "Choices and preferences: Evidence from implicit choices and response times." *Journal of Experimental Social Psychology* 48 (6):1336-1342. doi: http://dx.doi.org/10.1016/j.jesp.2012.07.004.

Balcombe, Kelvin, Iain Fraser, and Eugene McSorley. 2015. "Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments." *Journal of Applied Econometrics* 30 (3):447-467. doi: 10.1002/jae.2383.

Banks, James, Cormac O'Dea, and Zoë Oldfield. 2010. "Cognitive Function, Numeracy and Retirement Saving Trajectories*." *The Economic Journal* 120 (548):F381-F410. doi: 10.1111/j.1468-0297.2010.02395.x.

Bello, Muhammad, and Awudu Abdulai. 2016. "Impact of Ex-Ante Hypothetical Bias Mitigation Methods on Attribute Non-Attendance in Choice Experiments." *American Journal of Agricultural Economics*. doi: 10.1093/ajae/aav098.

Brock, William A., and Steven N. Durlauf. 2001. "Discrete Choice with Social Interactions." *The Review of Economic Studies* 68 (2):235-260. doi: 10.1111/1467-937X.00168.

Brown, Morton B., and Alan B. Forsythe. 1974. "Robust Tests for the Equality of Variances." *Journal of the American Statistical Association* 69 (346):364-367. doi: 10.2307/2285659.

Burkell, Jacquelyn. 2004. "What are the chances? Evaluating risk and benefit information in consumer health materials." *Journal of the Medical Library Association* 92 (2):200-208.

Campitelli, Guillermo, and Martín Labollita. 2010. "Correlations of cognitive reflection with judgments and choices." *Judgment and Decision Making* 5 (3):182.

Chen, Haipeng, Howard Marmorstein, Michael Tsiros, and Akshay R. Rao. 2012. "When More Is Less: The Impact of Base Value Neglect on Consumer Preferences for Bonus Packs over Price Discounts." *Journal of Marketing* 76 (4):64-77. doi: 10.1509/jm.10.0443.

ChoiceMetrics. 2014. NGENE v1.1.2 User Manual & Referene Guide. Sydney, Australia: Choice Metrics.

Cummings, Ronald G., and Laura O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *The American Economic Review* 89 (3):649-665.

de-Magistris, Tiziana, Azucena Gracia, and Jr Rodolfo M. Nayga. 2013. "On the Use of Honesty Priming Tasks to Mitigate Hypothetical Bias in Choice Experiments." *American Journal of Agricultural Economics* 95 (5):1136-1154. doi: 10.1093/ajae/aat052.

de Bekker-Grob, Esther W., Mandy Ryan, and Karen Gerard. 2012. "Discrete choice experiments in health economics: a review of the literature." *Health Economics* 21 (2):145-172. doi: 10.1002/hec.1697.

Duchowski, Andrew T. 2003. *Eye Tracking Methodology: Theory and Practice*: London : Springer London, 2003.

Dunning, David, Chip Heath, and Jerry M. Suls. 2004. "Flawed Self-Assessment:Implications for Health, Education, and the Workplace." *Psychological Science in the Public Interest* 5 (3):69-106. doi: 10.1111/j.1529-1006.2004.00018.x.

Einhäuser, Wolfgang, James Stout, Christof Koch, and Olivia Carter. 2008. "Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry." *Proceedings of the National Academy of Sciences* 105 (5):1704-1709. doi: 10.1073/pnas.0707727105.

Fagerlin, Angela, Brian J. Zikmund-Fisher, Peter A. Ubel, Aleksandra Jankovic, Holly A. Derry, and Dylan M. Smith. 2007. "Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale." *Medical Decision Making* 27 (5):672-680. doi: 10.1177/0272989x07304449.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *The Journal of Economic Perspectives* 19 (4):25-42. doi: 10.1257/089533005775196732.

Grebitus, Carola, Jayson L. Lusk, and Rodolfo M. Nayga. 2013. "Explaining differences in real and hypothetical experimental auctions and choice experiments with personality." *Journal of Economic Psychology* 36 (Supplement C):11-26. doi: https://doi.org/10.1016/j.joep.2013.02.004.

Grebitus, Carola, Jutta Roosen, and Carolin Claudia Seitz. 2015. Visual Attention and Choice: A Behavioral Economics Perspective on Food Decisions. In *Journal of Agricultural & Food Industrial Organization*.

Harrison, Glenn W., and E. Elisabet Rutström. 2008. "Chapter 81 Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods." In *Handbook of Experimental Economics Results*, edited by R. Plott Charles and L. Smith Vernon, 752-767. Elsevier.

Hensher, David A. 2006. "How do respondents process stated choice experiments? Attribute consideration under varying information load." *Journal of Applied Econometrics* 21 (6):861-878.

Hensher, David A., John Rose, and William H. Greene. 2005. "The implications on willingness to pay of respondents ignoring specific attributes." *Transportation* 32 (3):203-222. doi: 10.1007/s11116-004-7613-8.

Hensher, David A., John M. Rose, and William H. Greene. 2012. "Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design." *Transportation* 39 (2):235-245. doi: 10.1007/s11116-011-9347-8.

Hess, Stephane, David A. Hensher, and Andrew Daly. 2012. "Not bored yet – Revisiting respondent fatigue in stated choice experiments." *Transportation Research Part A: Policy and Practice* 46 (3):626-644. doi: http://dx.doi.org/10.1016/j.tra.2011.11.008.

Hole, Arne Risa, Julie Riise Kolstad, and Dorte Gyrd-Hansen. 2013. "Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour." *Journal of Economic Behavior & Organization* 96:21-31. doi: http://dx.doi.org/10.1016/j.jebo.2013.09.009.

Hoppe, Eva I., and David J. Kusterer. 2011. "Behavioral biases and cognitive reflection." *Economics Letters* 110 (2):97-100. doi: https://doi.org/10.1016/j.econlet.2010.11.015.

Hoyos, David. 2010. "The state of the art of environmental valuation with discrete choice experiments." *Ecological Economics* 69 (8):1595-1603. doi: https://doi.org/10.1016/j.ecolecon.2010.04.011.

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

Kahan, Dan M. 2013. "Ideology, motivated reasoning, and cognitive reflection." *Judgment and Decision Making* 8 (4):407-424.

Khachatryan, Hayk, Alicia Rihn, Benjamin Campbell, Chengyan Yue, Charles Hall, and Bridget Behe. 2017. "Visual Attention to Eco-Labels Predicts Consumer Preferences for Pollinator Friendly Plants." *Sustainability* 9 (10):1743.

Kløjgaard, Mirja Elisabeth, Mickael Bech, and Rikke Søgaard. 2012. "Designing a Stated Choice Experiment: The Value of a Qualitative Process." *Journal of Choice Modelling* 5 (2):1-18. doi: http://dx.doi.org/10.1016/S1755-5345(13)70050-2.

Krajbich, Ian, Carrie Armel, and Antonio Rangel. 2010. "Visual fixations and the computation and comparison of value in simple choice." *Nat Neurosci* 13 (10):1292-1298. doi: http://www.nature.com/neuro/journal/v13/n10/abs/nn.2635.html#supplementary-information.

Lipkus, Isaac M., Greg Samsa, and Barbara K. Rimer. 2001. "General Performance on a Numeracy Scale among Highly Educated Samples." *Medical Decision Making* 21 (1):37-44. doi: 10.1177/0272989x0102100105.

List, John A. 2001. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *The American Economic Review* 91 (5):1498-1507.

List, JohnA, and CraigA Gallet. 2001. "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics* 20 (3):241-254. doi: 10.1023/A:1012791822804.

Loo, Ellen J. Van, Rodolfo M. Nayga Jr., Danny Campbell, Han-Seok Seo, and Wim Verbeke. forthcoming. "Using Eye Tracking to Account for Attribute Non-Attendance in Choice Experiments." *European Review of Agricultural Economics*.

Louviere, Jordan J. 2006. "What You Don't Know Might Hurt You: Some Unresolved Issues in the Design and Analysis of Discrete Choice Experiments." *Environmental and Resource Economics* 34:173-188.

Luchini, S., and V. Watson. 2014. "Are choice experiments reliable? Evidence from the lab." *Economics Letters* 124 (1):9-13. doi: http://dx.doi.org/10.1016/j.econlet.2014.04.005.

Lusardi, Annamaria. 2012. Numeracy, financial literacy, and financial decision-making. National Bureau of Economic Research.

Lusk, Jayson L. 2003. "Effects of Cheap Talk on Consumer Willingness-to-Pay for Golden Rice." *American Journal of Agricultural Economics* 85 (4):840-856. doi: 10.1111/1467-8276.00492.

Malone, Trey, and Jayson L. Lusk. 2018. "Consequences of Participant Inattention with an Application to Carbon Taxes for Meat Products." *Ecological Economics* 145 (Supplement C):218-230. doi: https://doi.org/10.1016/j.ecolecon.2017.09.010.

Maughan, Lizzie, Sergei Gutnikov, and Rob Stevens. 2007. "Like more, look more. Look more, like more: The evidence from eye-tracking." *Journal of Brand Management* 14 (4):335-342. doi: 10.1057/palgrave.bm.2550074.

McFadden, Daniel. 1974. "The measurement of urban travel demand." *Journal of Public Economics* 3 (4):303-328. doi: http://dx.doi.org/10.1016/0047-2727(74)90003-6.

Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, and Darryl Weatherhead. 2005. "A Meta-analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental and Resource Economics* 30 (3):313-325. doi: 10.1007/s10640-004-3332-z.

Peters, Ellen, Daniel Västfjäll, Paul Slovic, C.K. Mertz, Ketti Mazzocco, and Stephan Dickert. 2006. "Numeracy and Decision Making." *Psychological Science* 17 (5):407-413. doi: 10.1111/j.1467-9280.2006.01720.x.

Rasch, Carsten, Jordan J Louviere, and Thorsten Teichert. 2015. "Using facial EMG and eye tracking to study integral affect in discrete choice experiments." *Journal of choice modelling* 14:32-47.

Robinson, Peter. 1998. "Literacy, numeracy and economic performance." *New Political Economy* 3 (1):143-149. doi: 10.1080/13563469808406341.

Sælensminde, Kjartan. 1998. "The impact of choice inconsistencies on the valuation of travel time in stated choice studies." First World Congress of Environmental and Resource Economists, Venice June.

Scarpa, Riccardo, Raffaele Zanoli, Viola Bruschi, and Simona Naspetti. 2013. "Inferred and Stated Attribute Non-attendance in Food Choice Experiments." *American Journal of Agricultural Economics* 95 (1):165-180. doi: 10.1093/ajae/aas073.

Schwartz, L. M., S. Woloshin, W. C. Black, and H. Welch. 1997. "THe role of numeracy in understanding the benefit of screening mammography." *Annals of Internal Medicine* 127 (11):966-972. doi: 10.7326/0003-4819-127-11-199712010-00003.

Silva, Andres, Rodolfo M. Nayga, Benjamin L. Campbell, and John L. Park. 2011. "Revisiting Cheap Talk with New Evidence from a Field Experiment." *Journal of Agricultural and Resource Economics* 36 (2):280-291.

Vossler, Christian A., Maurice Doyon, and Daniel Rondeau. 2012. "Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments." *American Economic Journal: Microeconomics* 4 (4):145-171. doi: 10.1257/mic.4.4.145.

Wang, Joseph Tao-yi, Michael Spezio, and Colin F. Camerer. 2010. "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games." *The American Economic Review* 100 (3):984-1007.

Weller, Joshua A., Nathan F. Dieckmann, Martin Tusler, C. K. Mertz, William J. Burns, and Ellen Peters. 2013. "Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach." *Journal of Behavioral Decision Making* 26 (2):198-212. doi: 10.1002/bdm.1751.