Judgment and Decision Making, Vol. 16, No. 6, November 2021, pp. 1464–1484

Incentivized and non-incentivized liking ratings outperform willingness-to-pay in predicting choice

Joshua Hascher* Nitisha Desai† Ian Krajbich‡

Abstract

A core principle in decision science is that people choose according to their subjective values. These values are often measured using unincentivized scales with arbitrary units (e.g., from 0 to 10) or using incentivized willingness-to-pay (WTP) with dollars and cents. What is unclear is whether using WTP actually improves choice predictions. In two experiments, we compare the effects of three different subjective valuation procedures: an unincentivized rating scale, the same scale with incentives, and incentivized WTP. We use these subjective values to predict behavior in a subsequent binary food-choice task. The unincentivized rating task performed better than the incentivized WTP task and no worse than the incentivized rating task. These findings challenge the view that subjective valuation tasks need to be incentivized. At least for low-stakes decisions, commonly used measures such as WTP may reduce predictive power.

Keywords: decision-making, valuation, incentivization, choice consistency, willingness-to-pay, Becker DeGroot Marschak auction

1 Introduction

Although experiments throughout the social sciences often rely on incentives, their importance in any given domain is debatable and not necessarily agreed upon across fields. For

*Department of Psychology, and Department of Economics, The Ohio State University. https://orcid.org/0000-0002-8283-8338.

†Department of Psychology, The Ohio State University. https://orcid.org/0000-0002-4595-8963.

‡Department of Psychology, and Department of Economics, The Ohio State University. Email: krajbich.1@osu.edu. https://orcid.org/0000-0001-6618-5675.

The pre-registration for Experiment 1 is available at https://aspredicted.org/blind.php?x=4za2dh.

The pre-registration for Experiment 2 is available at https://aspredicted.org/blind.php?x=w7zm9w.

I.K. gratefully acknowledges NSF Career Award 1554837 and the Cattell Sabbatical Fund, and J.H. gratefully acknowledges the OSU Decision Sciences Collaborative.

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

example, experimental economics views incentive structure as an essential tool for experimenters to obtain optimal response behavior from their subjects (Smith, 1982). There is a consensus among economists that salient incentives (incentives that depend on the decisions made by the subject, hereafter just "incentives") generally improve performance (Hertwig & Ortmann, 2001). Therefore, incentives are generally considered to be a requirement for publication in economics journals (Roth, 1995; Loewenstein, 1999; Camerer & Hogarth, 1999; Cox & Sadiraj, 2019). However, in the field of judgment and decision making (JDM) incentives are not always used (Hertwig & Ortmann, 2001). The more recent field of neuroeconomics/decision-neuroscience falls somewhere in the middle, where incentives are common but not required. Studies will often use hybrid approaches, e.g., having subjects make real Yes/No purchase decisions on a 4-point scale: {Strong Yes, Weak Yes, Weak No, Strong No}, where the incentives for "Strong Yes" and "Weak Yes" (as well as "Strong No" and "Weak No") are the same (Hare et al., 2009).

One justification for incentives is that decisions are effortful, and effort is costly (Hull, 1943; Kool et al. 2010). Therefore, experimenters must provide some benefit for accurate decisions in order to outweigh the effort cost (Smith & Walker, 1993). In a similar vein, subjects may not want to share private information if doing so will not benefit them in some way or if they are uncertain about exactly how their information will be used (Acquisti et al., 2015).

On the other hand, one justification for not needing incentives is that they may not be necessary to obtain effort. Subjects may feel obliged to take experiments seriously simply out of social obligation, or because they are receiving participation credit or payment (Camerer, 1995; Dawes, 1996; Gneezy & Rustichini, 2000a). Research has shown that people will reciprocate payments with effort, even if the payment is not conditional on the effort (Fehr et al., 1993). Additionally, the costs of making accurate decisions may be minimal and there may be some benefits. For instance, we know that people often derive utility from sharing information about themselves (Tamir et al., 2015; Tamir & Mitchell, 2012). Even economists have acknowledged that when tasks are enjoyable, incentives may not be important (Cox & Sadiraj, 2019).

There are also arguments against the use of incentives: they may sometimes cause more problems than they solve. For instance, providing monetary incentives may crowd out intrinsic motivation and backfire if the incentives are too small (Lepper et al., 1973; Gneezy & Rustichini, 2000b). On the other hand, introducing unrealistically large incentives could lead to excessive motivation, accentuating otherwise small biases (Arkes, 1991; Van Wallendael & Guignard, 1992; Beeler & Hunton, 1997). Finally, using incentives in some morally charged settings may violate social norms (Tetlock, 2003).

The empirical evidence that incentives matter is mixed and seems to depend on the domain (Jenkins et al., 1998; Camerer & Hogarth, 1999). When behavior can be evaluated objectively, incentives do seem to generally improve performance, though there are several examples where they do not make a difference (Hertwig & Ortmann, 2001). In social

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

decisions, incentives do seem to be important, increasing spiteful behavior and reducing egalitarian behavior (Bühren & Kundt, 2015), though they do not appear to impact social discounting (Locey et al., 2011). Likewise, incentives seem to be important in risky decisions, increasing risk aversion as stakes are raised (Holt & Laury, 2002). Meanwhile, incentives appear to have little effect on temporal discounting (Madden et al., 2004; Johnson & Bickel, 2002). Additionally, brain-imaging researchers have argued that subjects use one valuation system to make consumer decisions regardless of incentivization (Kang et al., 2011).

In this paper, we are specifically interested in the need for incentives in subjective valuation tasks. Much of the literature in decision-making is interested in how well individuals' choices align with the subjective value/utility they have for each option (e.g., Milosavljevic et al., 2010; Krajbich et al., 2010; Folke et al., 2016; Polanía et al., 2014; Pisauro et al., 2017; Philiastides & Ratcliff, 2013). That is, given a person's subjective values, how often does their decision process result in choosing the higher-value option? The answer to that question is undoubtedly influenced by how well we measure subjective value (Polanía et al. 2019).

The general structure of these experiments is to first elicit subjective-value ratings for a large set of items, then to study choices between pairs of those same items. The choice phase of these experiments is typically incentivized: subjects receive the outcome of one of their choices at the end of the study. However, the subjective-rating phase of these experiments is sometimes incentivized, sometimes not.

Many experiments use simple rating scales, for example a liking scale from –10 to +10, with 0 representing indifference (e.g., Krajbich et al., 2010; Krajbich & Rangel, 2011; Polanía et al., 2014; Polanía et al., 2019; Lim et al., 2011; Lebreton et al., 2009; Reutskaja et al., 2011; Litt et al., 2011; Towal et al., 2013). Although these types of rating procedures are easy to implement and easy for subjects to understand, they lack incentives to assign ratings carefully. This could cause subjects to put forth less effort and therefore report less precise ratings.

Other experiments elicit subjective values using the Becker-Degroot-Marschak (BDM) method (Becker et al., 1964) (e.g., Plassmann et al., 2007, 2010; Hare et al., 2008; De Martino et al., 2013; Kang et al., 2011; Linder et al., 2010; Jenison et al., 2011; Louie et al., 2013). In the BDM method, the subject assigns a monetary value to an item, known as their "willingness-to-pay" (WTP). The item is then assigned a random price by the experimenter. Then, if the subject's WTP is at or above the price, they receive the item and pay the price. Otherwise, they do not receive the item and they keep their money.

A major benefit of the BDM method is that it incentivizes people to truthfully report their WTP, which is taken as their subjective value. These values have the advantage of being on a scale with real units (i.e., dollars and cents), allowing us to predict how the items would fare against other goods outside of the experiment. This is not the case with ratings on arbitrary scales.

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

On the other hand, the BDM has some potential problems. As Harrison (1992) demonstrated, subjects are only weakly incentivized around their true WTP for each item: slight deviations are not very costly because subjects' bids affect only whether they will accept a random price. It has also been shown, both theoretically and experimentally, that the distribution of possible prices can influence how people bid (Horowitz, 2006; Tymula et al., 2016). In addition, the procedure itself may be confusing to subjects (Cason & Plott, 2014). They may believe that they should behave as if they were in an auction (first-price) and try to keep their bids low. Or perhaps some subjects assign what they believe to be the market value for an item instead of relying on their own subjective valuation (Thaler, 1985). Finally, it is difficult to assess negative values (i.e., aversive options) without extending WTP into the negative range, which is even more complicated to explain and requires the ability to inflict aversive outcomes on experimental subjects. But without extending WTP into the negative range, it is unclear whether a WTP of $0 means that a subject would take the item if it were free (indifference to receiving the item) or that they would rather receive nothing than receive the item (aversion to receiving the item) (Krajbich et al., 2012).

Here, we present a method that avoids the main drawbacks of the subjective rating and BDM procedures, combining the simplicity of the rating scale with the incentive-compatibility of the BDM. In this method, individuals assign subjective values to the items just as they would on a standard liking scale. To incentivize the ratings, they are told that at the end of the experiment two items will be selected at random and they will receive the higher rated item. This method is incentive compatible because subjects need to rate the items truthfully to ensure that they will receive their preferred item from the randomly selected pair. This method is similar in spirit to a method used to elicit cash equivalents for lotteries (Goldstein & Einhorn, 1987; Tversky et al., 1990).

In this pre-registered study (https://aspredicted.org/blind.php?x=4za2dh, https://aspredicted.org/blind.php?x=w7zm9w), we set out to test these three subjective-value procedures against each other to determine which method should be preferred. To do so, we compared each method's accuracy in predicting subjects' binary choices between food items in two separate experiments. In both, we employed a between-subjects design in which subjects were randomly assigned to one of the three rating procedures. Each subject rated a series of snack foods, and then made a series of choices between those same snack foods. We initially hypothesized that the incentivized rating task would perform best out of the three conditions, because it has the advantage of being incentivized but avoids the complications associated with WTP. We also hypothesized that both subjective-rating tasks would have fewer ties in the ratings, specifically fewer items rated at 0. We thought this might be the case since people might like some foods but not be willing to spend money on them (and not vice-versa). In the second experiment we hypothesized that the WTP condition would perform worse than the other two conditions (based on our first experiment).

To preview the results, we found that the incentivized rating procedure performed as well as, but not better than the unincentivized rating procedure, while the WTP procedure

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

performed worse than both.

2 Method

2.1 Subjects

In Experiment 1, subjects were recruited from a participant database in the Ohio State University Experimental Economics Laboratory. Using a power analysis on previously collected data, we determined that we would need 60 subjects for each condition in order to detect a 4% difference in accuracy between conditions, which we would consider meaningful support for that technique. In total, 183 subjects participated in the study. One subject's data were removed from the choice analyses because they did not rate enough items at or above zero to generate 300 pairs for the decision task. By condition, we analyzed data from 60 subjects in the unincentivized rating condition, 60 subjects in the WTP condition, and 61 subjects in the food-payment condition.

Compensation for the subjects varied by condition. All subjects earned at least $12, while subjects in the WTP condition could earn up to $16. Subjects in any condition could also receive one food item. To increase motivation, subjects were instructed not to eat for three hours prior to the beginning of the study. Subjects took a median of approximately 26 minutes to complete the experiment.

In Experiment 2, subjects were recruited from the same database as in Experiment 1, this time for an online experiment. Using a power analysis based on the Experiment 1's data, we determined that we would need 65 subjects in each condition to detect the same effects with 80% power. In total, 223 subjects participated, and we had to exclude the data from 17 subjects because they did not rate enough items at or above zero. By condition, we analyzed data from 72 subjects in the unincentivized rating condition, 65 subjects in the WTP condition, and 69 subjects in the food-payment condition.

All subjects earned at least $6, while subjects in the WTP condition could earn up to $10. Subjects in any condition could also receive one food item; one in ten subjects were randomly selected to have one of their choices implemented and a food shipped to them. Subjects took a median of 16 minutes to complete the experiment.

2.2 Materials

In Experiment 1, all instructions and stimuli were presented to subjects on computers in the Ohio State University Experimental Economics Laboratory using MATLAB's (MathWorks, 2014) Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). One hundred forty-four snack foods such as chocolate, candy, chips, etc. comprised the stimulus set, similar to sets used in previous research (e.g., Smith & Krajbich, 2018; Krajbich, et al., 2010). Subjects viewed a slideshow of all 144 food items at the beginning of the experiment. During this slideshow, each food was displayed for 750 ms., one at a time.

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

Experiment 2 was conducted online. The experiment was created using the JavaScript jsPsych extension (de Leeuw, 2015). Eighty healthy and unhealthy snack foods (e.g., berries, nuts, chocolates) comprised this stimulus set. The foods and images were taken from the website Nuts.com, allowing us to ship foods to the subjects.

In both experiments, subjects used a mouse to complete the rating task and a keyboard to complete the choice task. Both experiments were also pre-registered (as noted in the footnote on page 1).

2.3 Valuation task

The first task for subjects was to evaluate the set of food items. Within an experiment, the subjective-value scale itself was identical across conditions (Fig. 1), but the instructions for using the scale varied. In all conditions, subjects were first given instructions on how to use the scale (with a practice example in Experiment 1). They were instructed to assign each food item a value between 0 and 4, precise to two decimal places. In Experiment 1, subjects first saw each item for 2 seconds and then moved the mouse to their desired point on a semi-circular scale and clicked the left mouse button to indicate their rating. Once a subject's mouse touched the scale, their current mouse position's value was shown just below the scale. Experiment 2 instead used a linear scale, the item was displayed above the scale, the value was not displayed, and subjects confirmed their choice by clicking a "Continue" button below the scale. In both experiments, subjects could also indicate their aversion to an item by clicking a "Would Not Eat" button above the scale (above the item in Experiment 2). To provide some landmarks, each integer was shown next to the scale.

2.3.1 Unincentivized rating method

In this condition, subjects rated how much they would like to eat/receive that food after the study. There were no incentives in this task; subjects were told beforehand that their reward for the experiment would be determined only by their decisions in the subsequent choice task. The abbreviated instructions for this task were as follows (see Supplemental Material for full instructions):

> Experiment 1: "You will now rate each individual food based on how much you would like to eat that food at the end of the experiment. You will rate each food on a scale from 0 to 4. 4 means you would really like to eat the food. 0 means you would neither like nor dislike to eat the food. You may also click "Would Not Eat" if you would not like to eat the food. . .
>
> You will now be making your real food ratings. If you have any questions at this time, please ask the experimenter for help. It is important for you to be as accurate as possible with your ratings."

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

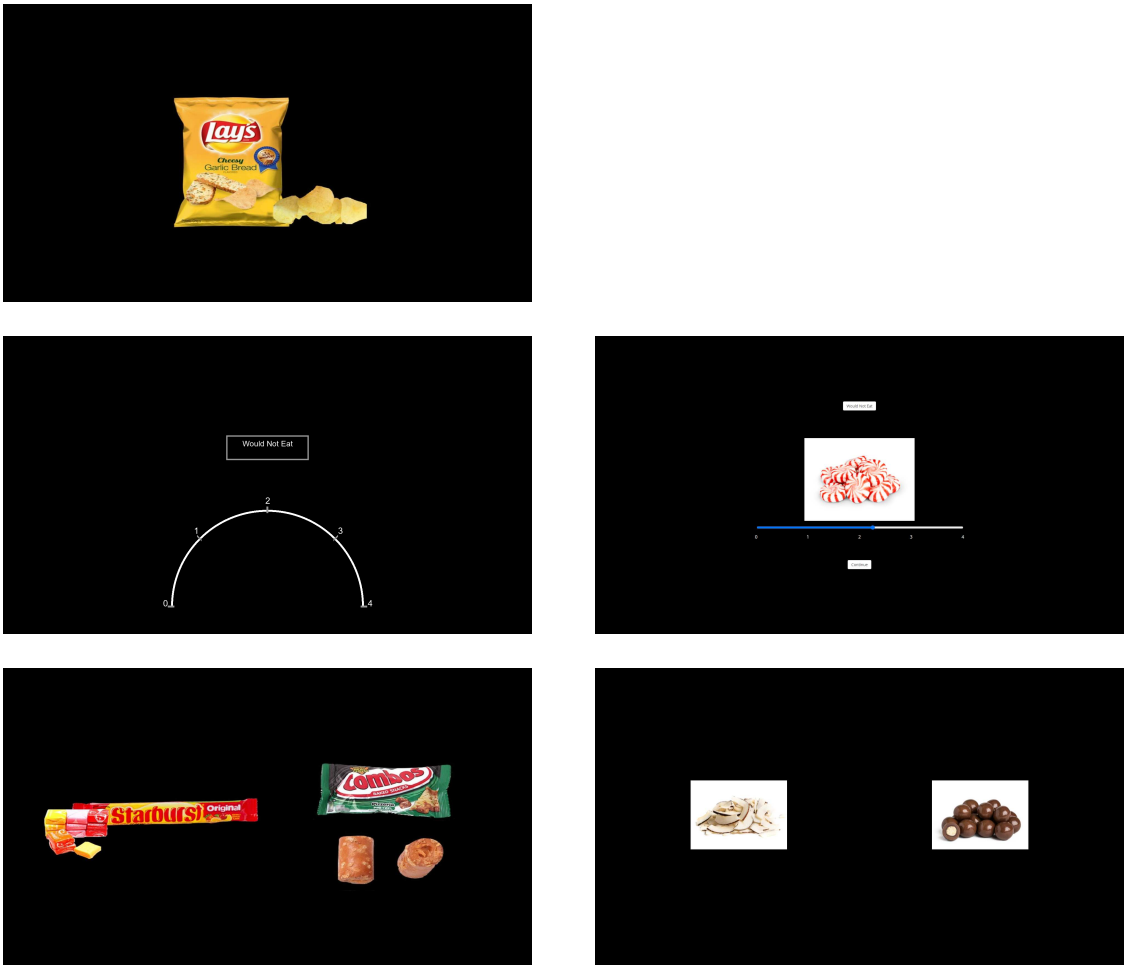

FIGURE 1: **Timeline of the experiments: (Left) Experiment 1, (Right) Experiment 2.** In the first phase, subjects provided subjective values for each food item. **(top row)** In Experiment 1, subjects first saw the item for 2 seconds. **(middle row)** In both experiments, subjects then provided their value for the items. This valuation screen was identical across the three conditions. **(bottom row)** In the second phase of the experiment, subjects made a series of choices between two randomly selected foods: 300 in Experiment 1 and 200 in Experiment 2.

> Experiment 2: "We are asking you to rate each food based on how much you would like to receive it. Rate each food on a scale from 0 to 4. 4 means that you would really like to eat it. 0 means you would neither like nor dislike to eat it. If you would not want to eat it, then click the "Would Not Eat" button.
>
> Use the mouse to click on the scale to indicate your rating. You will automatically progress to the next food."

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

2.3.2 Willingness-to-pay method

Subjects in the WTP condition reported how much they would be willing to pay to eat/receive each food at the end of the study through the BDM method. In Experiment 1, our instructions were based on recommendations for explaining the BDM procedure as clearly as possible (Healy, 2016; Healy, 2017) (Fig. 2). In Experiment 2, we used more standard BDM instructions, where subjects were told that they were to report how much they would be willing to pay for each food, given a $4 budget for each choice. Then, the reward scheme was explained as below.

Q#		Option A		Option B
1	Would you rather have:	Butterscotch	or	$4.00
2	Would you rather have:	Butterscotch	or	$3.99
3	Would you rather have:	Butterscotch	or	$3.98
⋮	⋮	⋮	⋮	⋮
400	Would you rather have:	Butterscotch	or	$0.01
401	Would you rather have:	Butterscotch	or	$0.00

FIGURE 2: **Experiment 1 WTP instructions.** To explain the BDM mechanism for incentivizing WTP, we used the following procedure: Subjects were told to think about each WTP decision as if it was a sequence of binary decisions between the food item or an amount of money equal to $0.01, $0.02, ..., $3.99, $4.00. They were told to report the point at which they would switch from receiving the money to receiving the food item.

This condition was incentivized; subjects were told that there was an equal chance that either the valuation task or the subsequent choice task would be used to determine their reward. If the valuation task was selected, the computer chose a random food item and price between $0 and $4. Depending on their WTP, the subject would then either receive the food item or receive money. In Experiment 1 this was described to subjects as a choice between the food and the money; in Experiment 2 it was described as setting a maximum price threshold. The abbreviated instructions for this task were as follows:

> Experiment 1: "You will now rate each individual food based on how much you would be willing to pay to have that food at the end of the study. Imagine you are going to rate this butterscotch candy. Now, I am going to ask you the following list of questions: (see Fig. 2)... In each question you pick either Option A (the butterscotch) or Option B (the money). If this food rating were randomly chosen for payment, I would randomly pick one question and pay you the option you chose on that one question. Each question is equally likely to be chosen for payment. Obviously you have no reason to lie on any question because if that question gets chosen for payment then you would end up with the option you like less.

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

I assume you are going to choose Option B in at least the first few questions, but at some point switch to choosing Option A. So, to save time, just tell me at which dollar value you would switch. I can then "fill out" your answers to all 401 questions based on your switch point (choosing Option B for all questions before your switch point, and Option A for all questions at or after your switch point). I will still draw one question randomly for payment. Again, if you lie about your true switch point you might end up getting paid an option that you like less. It is important to note that if this part of the study is randomly chosen for payment, then only one round will count (also randomly selected). This means that you should treat each food rating as if it is the only one.

If your reported switch point is $0.00, we will assume you would neither like nor dislike to eat the food. You may also click "Would Not Eat" if you would not like to eat that food. The catch is that if that food is chosen for payment, you will not receive any food.

Now, if this food were chosen at the end of the experiment, we would select a random question from the chart you saw earlier. Suppose the question offered you the choice between the butterscotch and $3.50. Because your switch point was $2.00, you decided to take the money for every value greater than $2.00. Therefore, in this example you would receive $3.50 instead of the butterscotch.

However, suppose instead that the random question offered you the choice between $0.50 and the butterscotch. Because your switch point is greater than that amount, you would receive the butterscotch.

You will now be making your real food ratings. If you have any questions at this time, please ask the experimenter for help."

Experiment 2: "We are asking you to report how much you would be willing to pay to receive each food, in the form of a bid. Indicate your bid for each food on a scale from $0 to $4, or click the 'Would Not Eat' button. If this task is randomly selected to determine your reward, ONE of the foods will be randomly selected, along with a random price from $0 to $4 in $0.01 increments.

If your bid is greater than or equal to the random price, you will get the food for that price. If your bid is less than the random price, you will NOT get the food, and will not have to pay anything. These rules ensure that it is in your best interest to bid your true willingness to pay for each food, since you cannot affect the price of the food, you can only decide what prices are acceptable to you. To help cover this potential cost, you will receive an additional $4 if this task is selected for payment. In that case you could earn between $6 and $10, depending on your bid and on the price.

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

> Use the mouse to click on the scale to indicate your rating. You will automatically progress to the next food.

2.3.3 Food-payment method

Subjects in the food-payment condition were instructed to use the value scale in the same manner as in the unincentivized rating condition. The key difference was that this condition was incentivized; subjects were told that there would be an equal chance of receiving a reward from the valuation task or the choice task. If the valuation task was selected, two random foods were drawn. The subject received whichever food they rated higher. If the foods had the same rating, the choice was made randomly. If both foods were rated "Would Not Eat," the subject received no food reward. The abbreviated instructions for this task were as follows:

> Experiment 1: "You will now rate each individual food based on how much you would like to eat that food at the end of the experiment. You will rate each food on a scale from 0 to 4. 4 means you would really like to eat the food. 0 means you would neither like nor dislike to eat the food. You may also click "Would Not Eat" if you would not like to eat the food.
>
> At the end of the study, if this rating task is selected for payment, a random pair of these foods will be selected, and you will receive the food that you rated higher. If there is a tie, one of the foods will be chosen at random. If you click "Would Not Eat" for both foods, you will not receive any food.
>
> You will now be making your real food ratings. If you have any questions at this time, please ask the experimenter for help."
>
> Experiment 2: "We are asking you to rate each food based on how much you would like to receive it. Rate each food on a scale from 0 to 4. 4 means that you would really like to eat it. 0 means you would neither like nor dislike to eat it. If you would not want to eat it, then click the "Would Not Eat" button.
>
> If this task is randomly selected to determine your reward, a random pair of these foods will be selected and you will receive the food that you rated higher. If there is a tie, one of the foods will be chosen at random. If you selected "Would Not Eat" for both foods, you will not receive any food.
>
> Use the mouse to click on the scale to indicate your rating. You will automatically progress to the next food."

2.4 Choice task

After the valuation task, subjects made binary choices between foods that they had rated at or above zero. Subjects made 300 choices in Experiment 1 and 200 choices in Experiment

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

2. They were given the opportunity to take a short break every 100 trials. In line with the valuation task, subjects were told to choose the food that they would prefer to eat at the end of the study (Experiment 1) or have shipped to them after the study (Experiment 2). The choice task itself did not vary across conditions. The abbreviated instructions for this task were as follows:

> Experiment 1: "You will now be making a series of choices between pairs of foods. Remember, one of these choices may/will be used to determine your reward. Use the left and right arrow keys to select which food you would prefer to eat at the end of the study."
>
> Experiment 2: "In this part of the study, you will see two foods on the screen. You have to choose which food you would prefer to eat. To select the left food, press the F key. To select the right food, press the J key. After each choice, stare at the white cross at the center of the screen. When you are ready, press the spacebar to begin with a couple of practice rounds.
>
> Now you can move on to the real choices. Imagine that you are in a shop and you are choosing between the two foods presented to you. Remember, the food you choose in one of the rounds may be shipped to you after the study."

3 Results

For Experiment 1, the average (se) value in each condition after pooling ratings by subject was: 2.12 (0.07) in the unincentivized condition, \$1.09 (\$0.10) in the WTP condition, and 2.02 (0.10) in the food-payment condition. For Experiment 2, the average values (in the same order) were: 2.14 (0.06), \$1.69 (\$0.09), and 2.15 (0.09) (Fig. 3).

Our key variable of interest was choice accuracy within each condition. We defined a "correct" decision in the binary-choice task as one which was in accordance with the subject's values. In other words, a choice was labeled as correct if the subject chose the item that they valued higher. If both items were rated the same for any choice, it was randomly labeled as correct or incorrect.

We first computed the mean accuracy level for each subject then pooled subjects by condition (Fig. 4a). In Experiment 1, the unincentivized condition's mean (se) accuracy was 78.4% (1.1%), the WTP condition's accuracy was 71.8% (1.6%), and the food-payment condition's accuracy was 76.1% (1.6%). For Experiment 2, the mean (se) accuracies were: 73.6% (0.9%), 71.3% (1.3%), and 73.2% (1.3%). We tested for a difference between the conditions with a one-way ANOVA (with the unincentivized condition as the baseline group), which revealed that the mean accuracy level was not equivalent across conditions for Experiment 1 ($F_{(2,178)} = 8.82$, $p < 0.001$). Additionally, using Tukey's Honestly Significant Difference post-hoc test to maintain a familywise α of 0.05, we found that the WTP condition had significantly lower accuracy than the unincentivized condition ($q = 5.87$, p

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

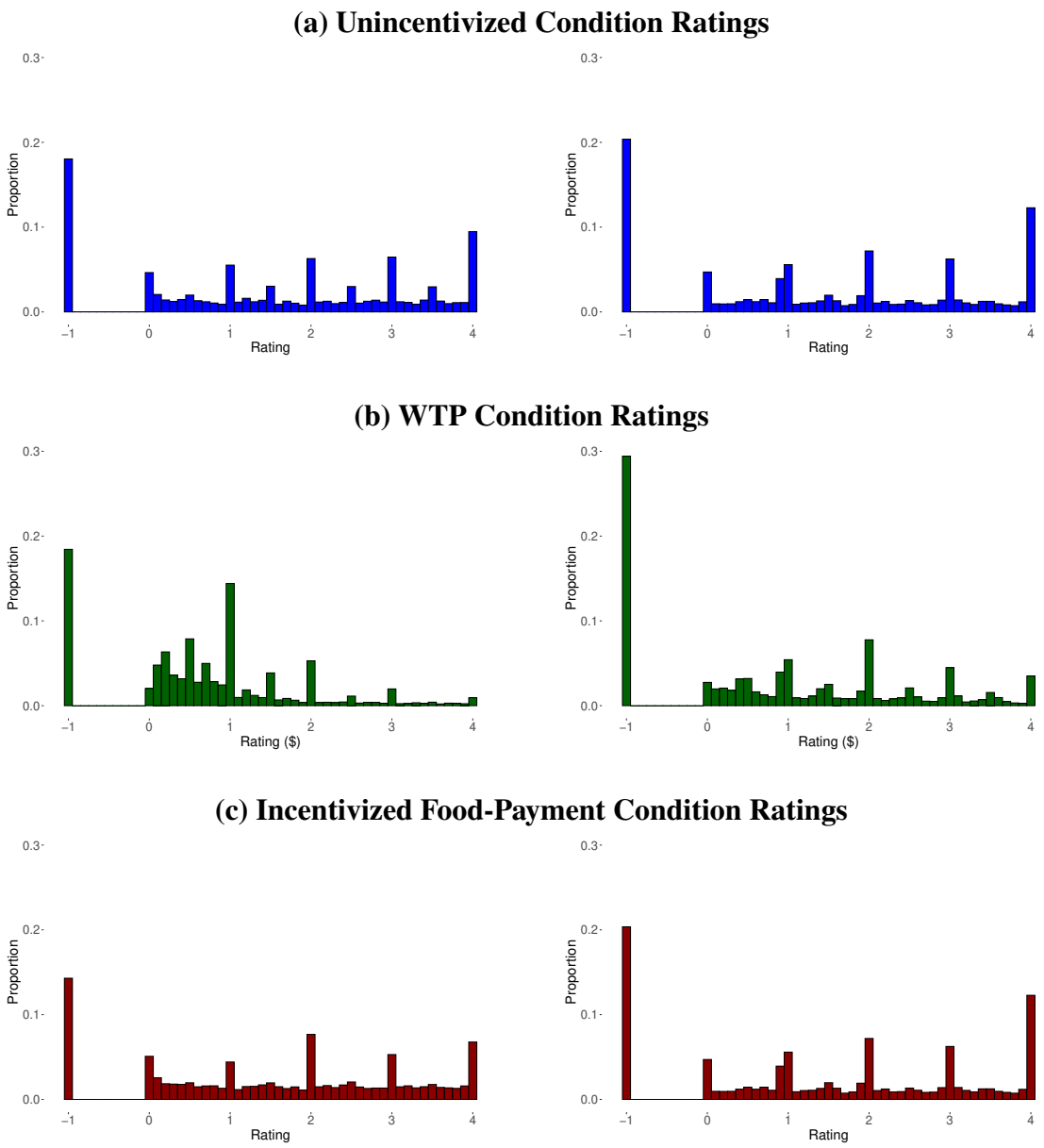

FIGURE 3: **Subjective-value histograms by experiment: (Left) Experiment 1, (Right) Experiment 2.** The pooled distribution of subjective-values in **(a)** the unincentivized rating condition, **(b)** the WTP condition, and **(c)** the food-payment condition. A value of -1 corresponds to cases where subjects chose the "Would Not Eat" option.

< 0.001) as well as the food-payment condition (q = 3.81, p = 0.02). The unincentivized and food-payment conditions were not significantly different from each other (q = 2.05, p = 0.3). In Experiment 2, the one-way ANOVA revealed no significant difference across the conditions ($F_{(2,203)}$ = 2.261, p = 0.1). Nevertheless, per our hypotheses, we tested whether WTP was worse than the two rating conditions using one-tailed t-tests. WTP was worse than the unincentivized ratings (t = 1.68, p = 0.048) but not significantly worse than the food-payment ratings (t = 1.33, p = 0.09).

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

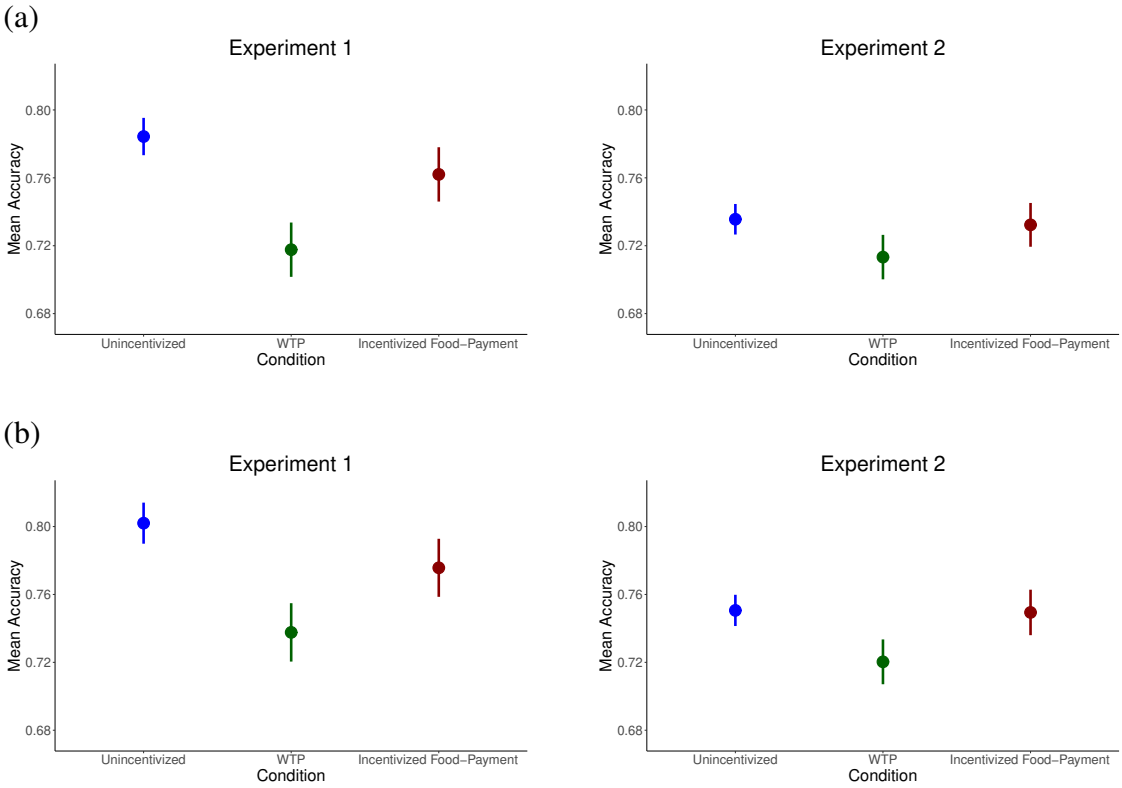

FIGURE 4: **Accuracy rates by condition: (a) with ties included, (b) without ties.** An accurate decision is one in which the subject chose the item with the higher subjective value from the valuation task. Error bars indicate standard errors, clustered by subject. **(b)** We first excluded all choices between equally valued items, then computed the accuracy measures.

Next, we recomputed accuracy rates in each condition, excluding all trials in which there was a tie in the ratings, and then re-ran our previous test (Fig. 4b). The new mean (se) accuracy rates for Experiment 1 were 80.2% (1.2%) for the unincentivized condition, 73.8% (1.7%) for the WTP condition, and 77.6% (1.7%) for the food-payment condition. The new rates for Experiment 2 were 75.1% (0.9%), 72.0% (1.3%), and 74.9% (1.3%), respectively. In Experiment 1, this again revealed a significant difference between the conditions ($F_{(2,178)} = 7.50$, $p < 0.001$). However, in the pairwise comparisons between conditions, the difference between the food-payment condition and the WTP condition was no longer significant at a familywise α of 0.05 ($q = 3.23$, $p = 0.06$). The results of the other two pairwise comparisons match those of the original analyses ($q = 5.46$, $p < 0.001$ between the unincentivized condition and the WTP condition; $q = 2.23$, $p = 0.26$ between the unincentivized condition and the food-payment condition). In Experiment 2, the one-way ANOVA now revealed a significant difference in the accuracy rates between conditions ($F = 3.20$, $p = 0.04$). Per our hypotheses, we tested whether WTP was worse than the two rating conditions using one-tailed t-tests. WTP was significantly worse than the unincentivized ratings ($t = -2.26$, $p = 0.03$) and significantly worse than the food-payment ratings ($t = -2.24$, $p = 0.03$).

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

To better understand how the subjective-rating conditions outperformed the WTP condition, we investigated several possibilities. We hypothesized that there might be more ties in the WTP condition. We computed a "ties" statistic in the same way as our accuracy test above and conducted a second ANOVA to compare the conditions. We failed to find a significant difference between the conditions in either experiment ($F_{(2,178)} = 0.78$, $p = 0.46$; $F_{(2,203)} = 2.62$, $p = 0.08$), and the direction of the difference, especially in the second experiment, was opposite to what we expected, with fewer ties in the WTP condition. We also tested our hypothesis that the WTP condition would have more subjective values of 0. We again failed to find a significant difference in either experiment ($F_{(2,178)} = 2.55$, $p = 0.08$; $F_{(2,203)} = 1.98$, $p = 0.14$), and again the difference was in the wrong direction, with fewer zeros in the WTP condition.

It is also possible that the conditions differed in the rates at which subjects assigned ratings of "Would Not Eat". Following the procedure above, we computed a subject-level proportion of items that were rated "Would Not Eat" and conducted a one-way ANOVA. For Experiment 1, we failed to find any significant differences ($F_{(2,178)} = 0.91$, $p = 0.41$), but for Experiment 2 we found that the averages were different ($F_{(2,203)} = 4.18$, $p = 0.02$), and that the WTP condition had significantly more ratings of "Would Not Eat" than both the unincentivized ($q = 3.52$, $p = 0.04$) and the incentivized food-payment ($q = 3.66$, $p = 0.03$) conditions.

We also performed an analysis to check for the possibility that monetary incentives might have crowded out intrinsic motivation to do the WTP task carefully. To do so, we compared subjects' median response times on the valuation task between conditions. Contrary to the idea that subjects were less motivated on the WTP task, they actually took more time on that task than on the other two valuation tasks ($F_{(2,178)} = 16.8$, $p < 0.001$; $q = 5.90$, $p < 0.001$ in comparison to the unincentivized ratings; and $q = 7.88$, $p < 0.001$ in comparison to the food-payment ratings). However, in Experiment 2 this difference was not significant ($F_{(2,203)} = 2.38$, $p = 0.1$), though the difference was in the same direction. Excluding "Would Not Eat" items did yield significant differences between WTP and the other two conditions ($F_{(2,203)} = 7.361$, $p < 0.001$; $q = 3.39$, $p = 0.02$ and $q = 5.32$, $p < 0.001$ for the comparisons with unincentivized and food-payment ratings respectively) and made the differences even larger and more significant in Experiment 1 ($F(2,178) = 25.5$, $p < 0.001$; $q = 7.30$, $p < 0.001$; $q = 9.70$, $p < 0.001$ in the same order as above). It is worth noting that these differences in median RT are quite large in magnitude: they yield a mean increase of 438 ms and 495 ms relative to the unincentivized task (which had means of 1476 ms and 3973 ms) in Experiments 1 and 2 respectively.

An inspection of Figure 3 suggests that one difference between conditions might be that values were less uniformly distributed in the WTP condition compared to the two other conditions. We confirmed this with Kolmogorov-Smirnov (KS) tests between the pooled value distributions in each condition and uniform distributions. The KS D-statistics were 0.11, 0.418, and 0.069 for the unincentivized, WTP, and food-payment conditions,

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

respectively (higher D = less uniform). This was also true in Experiment 2, though the difference was less extreme (D = 0.128, 0.166, and 0.145).

Additionally, we compared the average variance in values across the conditions. In Experiment 1 the average subject-level variances were 1.35, 0.452, and 1.31 in the unincentivized, WTP, and food-payment conditions, respectively. In Experiment 2 the subject-level variances were 1.43, 0.996, and 1.45. Thus, there was indeed substantially more variance within individuals in the rating conditions relative to the WTP condition. We can also examine these subject-level variances as a fraction of the pooled variance within each condition. In Experiment 1 the ratios of subject-level variance to total variance were 0.825, 0.568, and 0.835. In Experiment 2 these were 0.891, 0.748, and 0.868. So, the WTP condition indeed showed less variance within subjects relative to between subjects.

These analyses suggested to us that a likely explanation for the inferiority of WTP in our experiments is that subjects' WTPs are generally constrained to lie in a smaller range of the scale (and thus less uniform) resulting in more overlap of the noisy distributions of the items' values. This overlap leads to the appearance of more preference reversals (Polanía et al., 2019). We tested this idea by running correlations (pooling the unincentivized and food-payment conditions) between a subject's accuracy and the uniformity of their ratings (based on the KS D-statistic). Indeed, there were positive correlations between accuracy and uniformity in both Experiment 1 (Pearson's $r = 0.44$, $p < 0.001$) and Experiment 2 ($r = 0.31$, $p < 0.001$). Thus, given the less uniform distributions in the WTP condition, we would indeed expect lower accuracy in that condition.

We also tested this idea by expanding our definition of ties to include any pairs of items with subjective values within either 0.25 or 0.5 of each other. These "pseudo-tie" trials were then excluded before recomputing accuracies. Indeed, we found that excluding these trials substantially reduced the difference between conditions. Excluding pseudo-ties in Experiment 1, we found accuracies of 83.3%, 78.1%, and 81.1% (excluding ties within 0.25); and 84.9%, 82.4%, and 83.9% (excluding ties within 0.5). In the first case, WTP was no longer significantly worse than the food-payment condition ($q = 3.41$, $p = 0.046$; $q = 1.96$, $p = 0.36$, for unincentivized ratings and food-payment ratings excluding ties within 0.25; $q = 2.02$, $p = 0.34$; $q = 0.98$, $p = 0.77$, in the same order, excluding ties within 0.5). In Experiment 2, the accuracy differences between conditions did not appreciably change, with accuracies of 79.2%, 75.7%, and 78.4% (excluding ties within 0.25); and 80.7%, 78.2%, and 80.4% (excluding ties within 0.5). Again, in the last case WTP wss no longer significantly worse than the other two conditions ($q = 2.30$, $p = 0.24$ for unincentivized ratings and $q = 1.96$, $p = 0.36$ for food-payment ratings). Thus, we did find evidence that part of the problem with WTP is with the compression of the ratings.

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

4 Discussion

The purpose of this study was to test whether the method for eliciting subjective values influences how well those values predict choices between alternatives. We found that willingness-to-pay (WTP) made significantly worse predictions than both unincentivized ratings and incentivized ratings. Additionally, we found no significant difference between incentivized and unincentivized ratings; if anything, there is weak evidence that unincentivized ratings performed better. This was counter to our initial expectations that the incentivized rating system would perform the best, though in line with our hypotheses following the first experiment.

To better understand the difference between conditions, we explored the number of negatively rated items, the number of zeros, and the number of ties. There were no consistent differences between conditions on any of these measures, and the results remained largely unchanged when choices between equally valued items were removed from the accuracy analysis.

We did observe that the WTP values were substantially less uniformly distributed than the subjective ratings. Importantly, we established (using only the subjective rating conditions) that less uniformity is associated with lower accuracy. We also found that by excluding trials in which items are close in value, the difference between conditions decreases and becomes non-significant. These exploratory analyses suggest that the problems with WTP are that it induces value distributions that are more Gaussian than uniform and that use less of the value scale. The combination of these two factors leads to more cases where the values of the items are so close together that their inherent variability makes it difficult to form an accurate prediction. This is not a problem with WTP per se, or with subjects' use of the scales. It instead reflects that WTP and subjective ratings measure two different things; the former is a cardinal measure of economic value (and so should be Gaussian distributed), while the latter is an ordinal measure of utility (and so should be uniformly distributed). We did observe slightly more uniform WTPs in Experiment 2 than in Experiment 1, perhaps due to the different set of foods, the online setting, or the different instructions.

There are several additional explanations for why WTP might underperform compared to the other rating procedures. First, it could be that the incentives of the BDM mechanism are not strong enough. This seems unlikely given that WTP performed worse than the unincentivized rating task. Additionally, subjects on average took more time to complete the WTP valuation task than the other rating tasks. Second, the WTP procedure may be confusing. Although the level of confusion caused by the instructions would be difficult to measure, we did take special care to use instructions that were designed to maximize subject comprehension (Healy, 2016; Healy, 2017) (in Experiment 1), while also using standard instructions (in Experiment 2). Thus, our study is likely a lower bound on how badly confusion can affect WTP. Third, WTP may make it more difficult to account for inter-subject differences, which has been shown in the domain of public goods (Kahneman et al., 1993). For some, the scale may be too wide, with most of their ratings restricted to a small

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

section. For others, the scale may be too narrow, leading to ratings clustered at the edges. In contrast, a rating scale with arbitrary units allows the decision maker to spread their ratings out over the whole scale. We indeed found evidence for less within-subject variance in their WTPs than in their subjective ratings. Finally, WTP may cause subjects to partially focus on the market value of the goods rather than their own subjective values (Thaler, 1985). This would be consistent with some interpretations of prior work on preference reversals and joint vs. separate evaluations of alternatives, where context changes the relative weight of different dimensions on choice (Lichtenstein & Slovic 1971; Grether & Plott 1979; Hsee 1996).

There are also several reasons why the unincentivized ratings might perform as well as the incentivized food-payment ratings. Subjects may find it easiest to simply state their true values rather than to devise some other strategy. They may also find it intrinsically rewarding to give accurate information about themselves (Tamir et al., 2015; Tamir & Mitchell, 2012). Finally, people (particularly those who choose to participate in experiments) may feel obliged to put in the effort to do the ratings accurately (Fehr et al., 1993). All of these potential causes could diminish the importance of providing incentives when eliciting subjective values.

Overall, our results suggest that the use of incentives in low-stakes valuation tasks may not be necessary and may in fact be counterproductive. Here we see that incentivized food-payment ratings do not outperform unincentivized ratings, and that WTPs underperform both. Thus, if a researcher's goal is to predict choices between items in their experiment, they should consider using subjective ratings instead of WTP.

On the other hand, WTP does undoubtedly have some advantages over subjective rating scales. These values are on a scale with real units (i.e., dollars and cents), allowing us to predict purchase rates or to make comparisons with other items beyond the experimental context. This is not the case with arbitrary rating scales. For these reasons, researchers may still want to measure WTP, depending on their goals.

Despite the importance of eliciting subjective values, there is no standard method for doing so. Some experiments in psychology do not incentivize rating tasks at all, while incentive-compatible methods such as the BDM are routine in experimental economics. For those who simply want to maximize predictive power within their experimental context, our results indicate that unincentivized ratings should be sufficient. For those who prefer to play it safe and maintain incentivization, we recommend the incentivized rating procedure over the BDM mechanism, as it performs better and is simpler to explain.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514.

https://doi.org/10.1017/S1930297500008500 Published online by Cambridge University Press

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486–498.

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

Beeler, J. D., & Hunton, J. E. (1997). The influence of compensation method and disclosure level on information search strategy and escalation of commitment. *Journal of Behavioral Decision Making*, *10*(2), 77–91.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Bühren, C., & Kundt, T. C. (2015). Imagine being a nice guy: A note on hypothetical vs. incentivized social preferences. *Judgment and Decision Making*, *10*(2), 185–190.

Camerer, C. (1995). Individual decision making. *Handbook of experimental economics*.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(1), 7–42.

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

Cox, J. C., & Sadiraj, V. (2019). Incentives. *Handbook of Research Methods and Applications in Experimental Economics*. 9–27.

Dawes, R. M. (1996). The purpose of experiments: Ecological validity versus comparing hypotheses. *Behavioral and Brain Sciences*, *19*(1), 20–20.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *The Quarterly Journal of Economics*, *108*(2), 437–459.

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*(1), 1–8.

Gneezy, U., & Rustichini, A. (2000a). A fine is a price. *The Journal of Legal Studies*, *29*(1), 1–17.

Gneezy, U., & Rustichini, A. (2000b). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, *115*(3), 791–810.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomenon. *Psychological Review*, *94*(2), 236–254.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, *69*(4), 623–638.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*, *324*(5927), 646–648.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating

the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *Journal of Neuroscience*, *28*(22), 5623–5630.

Harrison, G. W. (1992). Theory and misbehavior of first-price auctions: Reply. *The American Economic Review*, *82*(5), 1426–1443.

Healy, P. J. (2016). *Explaining the BDM—Or any random binary choice elicitation mechanism—To Subjects*. mimeo, Ohio State University.

Healy, P. J. (2017). *Epistemic experiments: Utilities, beliefs, and irrational play*. mimeo, Ohio State University.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists?. *Behavioral and Brain Sciences*, *24*(3), 383–403.

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

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

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*(3), 247–257.

Hull, C. L. (1943). *Principles of Behavior* (Vol. 422). New York: Appleton-century-crofts.

Jenison, R. L., Rangel, A., Oya, H., Kawasaki, H., & Howard, M. A. (2011). Value encoding in single neurons in the human amygdala during decision making. *Journal of Neuroscience*, *31*(1), 331–338.

Jenkins Jr, G. D., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology*, *83*(5), 777–787.

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, *77*(2), 129–146.

Kahneman, D., Ritov, I., Jacowitz, K. E., & Grant, P. (1993). Stated willingness to pay for public goods: A psychological perspective. *Psychological Science*, *4*(5), 310–315.

Kang, M. J., Rangel, A., Camus, M., & Camerer, C. F. (2011). Hypothetical and real choice differentially activate common valuation areas. *Journal of Neuroscience*, *31*(2), 461–468.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, *3*, 193. 1–18.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the

relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857.

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron*, *64*(3), 431–439.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the" overjustification" hypothesis. *Journal of Personality and Social Psychology*, *28*(1), 129–137.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*(1), 46–55.

Lim, S. L., O'Doherty, J. P., & Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *Journal of Neuroscience*, *31*(37), 13214–13223.

Linder, N. S., Uhl, G., Fliessbach, K., Trautner, P., Elger, C. E., & Weber, B. (2010). Organic labeling influences food valuation and choice. *NeuroImage*, *53*(1), 215–220.

Litt, A., Plassmann, H., Shiv, B., & Rangel, A. (2011). Dissociating valuation and saliency signals during decision-making. *Cerebral Cortex*, *21*(1), 95–102.

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards. *Judgment and Decision Making*, *6*(6), 552–564.

Loewenstein, G. (1999). Experimental economics from the vantage-point of behavioural economics. *The Economic Journal*, *109*(453), F25-F34.

Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, *110*(15), 6139–6144.

Madden, G. J., Raiff, B. R., Lagorio, C. H., Begotka, A. M., Mueller, A. M., Hehli, D. J., & Wegener, A. A. (2004). Delay discounting of potentially real and hypothetical rewards: II. Between-and within-subject comparisons. *Experimental and Clinical Psychopharmacology*, *12*(4), 251–261.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making*, *5*(6), 437–449.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Philiastides, M. G., & Ratcliff, R. (2013). Influence of branding on preference-based decision making. *Psychological Science*, *24*(7), 1208–1215.

Pisauro, M. A., Fouragnan, E., Retzler, C., & Philiastides, M. G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nature Communications*, *8*(1), 1–9.

Plassmann, H., O'Doherty, J. P., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *Journal of Neuroscience, 27*(37),

9984–9988.

Plassmann, H., O'Doherty, J. P., & Rangel, A. (2010). Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. *Journal of Neuroscience, 30*(32), 10799–10808.

Polanía, R., Krajbich, I., Grueschow, M., & Ruff, C. C. (2014). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron*, *82*(3), 709–720.

Polanía, R., Woodford, M., & Ruff, C. C. (2019). Efficient coding of subjective value. *Nature Neuroscience*, *22*(1), 134–142.

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, *101*(2), 900–926.

Roth, A. E. (1995). 1. Introduction to experimental economics. In *The handbook of experimental economics* (pp. 1–110). Princeton University Press.

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General*, *147*(12), 1810–1826.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American Economic Review*, *72*(5), 923–955.

Smith, V. L., & Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, *31*(2), 245–261.

Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences*, *109*(21), 8038–8043.

Tamir, D. I., Zaki, J., & Mitchell, J. P. (2015). Informing others is associated with behavioral and neural signatures of value. *Journal of Experimental Psychology: General*, *144*(6), 1114–1123.

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, *7*(7), 320–324.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, *4*(3), 199–214.

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences*, *110*(40), E3858-E3867.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review, 80*(1), 204–217.

Tymula, A., Woelbert, E., & Glimcher, P. (2016). Flexible valuations for consumer goods as measured by the Becker–DeGroot–Marschak mechanism. *Journal of Neuroscience, Psychology, and Economics*, *9*(2), 65–77.

Van Wallendael, L. R., & Guignard, Y. (1992). Diagnosticity, confidence, and the need for information. *Journal of Behavioral Decision Making*, *5*(1), 25–37.